

CONVERGENCE OF SEQUENTIAL MONTE CARLO-BASED SAMPLING METHODS

JONATHAN H. HUGGINS AND DANIEL M. ROY

ABSTRACT. Originally designed for state-space models, Sequential Monte Carlo (SMC) methods are now routinely applied in the context of general-purpose Bayesian inference. Traditional analyses of SMC algorithms have focused on their application to estimating expectations with respect to intractable distributions such as those arising in Bayesian analysis. However, these algorithms can also be used to obtain approximate samples from a posterior distribution of interest. We investigate the asymptotic and non-asymptotic convergence rates of SMC from this sampling viewpoint. In particular, we study the expectation of the particle approximation that SMC produces as the number of particles tends to infinity. This “expected approximation” is equivalent to the law of a sample drawn from the SMC approximation. We give convergence rates of the Kullback-Leibler divergence between the target and the expected approximation. Our results apply to both deterministic and adaptive resampling schemes. In the adaptive setting, we introduce a novel notion of effective sample size, the ∞ -ESS, and show that controlling this quantity ensures stability of the SMC sampling algorithm. We also introduce an adaptive version of the conditional SMC proposal, which allows us to prove quantitative bounds for rates of convergence for adaptive versions of iterated conditional sequential Monte Carlo Markov chains and associated adaptive particle Gibbs samplers.

1. Introduction	2
2. Preliminaries	3
3. Summary of Results	10
4. Basic Results	13
5. SMC with Adaptive Resampling	20
6. Convergence in the Marginal and Predictive Cases	24
7. Particle MCMC	28
Appendix A. Additional Proofs	33
Acknowledgments	40
References	40

1. INTRODUCTION

Sequential Monte Carlo (SMC) methods are a widely-used class of algorithms for approximate inference [11–14, 16, 17]. In the context of Bayesian inference, SMC produces a particle approximation to the posterior distribution as well as an unbiased estimate of the marginal likelihood. Traditionally, particle approximations were intended to be used to estimate expectations such as the posterior probabilities of events or the expected values of parameters, and the analysis of SMC methods focused on this *operator* perspective, i.e., the analysis of how well a particle approximation approximates the expectation operator induced by the target distribution.

Increasingly, however, SMC methods are being used to produce approximate samples, usually in the inner loop of another approximate inference algorithm. A key example is the class of particle Markov chain Monte Carlo (PMCMC) methods, which aim to combine the best features of SMC and MCMC approaches by using SMC as a proposal mechanism for a Metropolis-Hastings (“particle MH”) or approximate Gibbs (“particle Gibbs”) sampler [1, 15]. Characterizing the efficiency of PMCMC methods is an active area of investigation [2–5, 18, 19].

When SMC methods are employed for sampling, convergence guarantees from the operator perspective are not appropriate. In this work, we take up the *measure* perspective on SMC, i.e., we characterize how well SMC-based methods can approximate the target distribution *as a measure* by assessing convergence in terms of total variation distance, KL divergence, and other measures of discrepancy between distributions.

Let π be a distribution of interest and let π^N denote an N -particle SMC approximation to π . In this work, the majority of our attention is devoted to investigating the mean of π^N , denoted $\bar{\pi}^N$, which can also be understood as the (marginal) distribution of a single sample drawn from the true posterior π^N . We use the Kullback-Leibler (KL) divergence from π to $\bar{\pi}^N$ to measure performance of the SMC sampler. We begin by giving convergence rates for sequential importance sampling and sampling importance resampling (also known as bootstrap filtering) (Section 4). We provide non-asymptotic bounds under minimal assumptions, which can then be used to obtain quantitative bounds under a number of different conditions. These quantitative bounds are obtained using recent results from [2] for bounding the expected value of the partition function estimator with respect to the conditional SMC kernel. We also obtain quantitative bounds for the KL divergence from π to $\bar{\pi}^N$ for adaptive SMC algorithms (Section 5). In the adaptive case our approach uses the α SMC framework [23], and thus applies to a large class of adaptive algorithms. We introduce a novel notion of effective sample size, ∞ -ESS, the control of which is sufficient to guarantee convergence of α SMC samplers.

We also propose a version of the conditional SMC kernel with adaptive resampling, which we call the conditional α SMC kernel (Section 7). Mirroring results from [2], we give conditions under which iterated conditional α SMC Markov chains are uniformly ergodic and the associated particle Gibbs sampler with adaptive resampling (the α PG sampler) is geometrically ergodic. As with the α SMC samplers, control of the ∞ -ESS is key to proving these ergodicity results.

2. PRELIMINARIES

After fixing some notation, we define the inference problem and present three SMC algorithms: sequential importance sampling, sampling importance resampling, and α SMC. We then introduce the conditional SMC kernel, as well as a novel adaptive version based on α SMC, which allows us to define adaptive versions of the iterated conditional SMC process and an associated particle Gibbs algorithm.

2.1. Notation. For a positive integer K , let $[K] \triangleq \{1, 2, \dots, K\}$. If x_i, \dots, x_j are elements of a sequence, write $x_{i,j} \triangleq \langle x_i, x_{i+1}, \dots, x_j \rangle$. We use the following conventions: $\sum_{\emptyset} = 0$, $\prod_{\emptyset} = 1$, and $0/0 = 0$.

Let (S, \mathcal{S}) , (S', \mathcal{S}') be measurable spaces. $K : S \times S' \rightarrow \mathbb{R}$ is a kernel if $K(\cdot, B)$ is a (S, \mathcal{S}) -measurable function for all $B \in \mathcal{S}'$ and $K(x, \cdot)$ is measure on (S', \mathcal{S}') for all $x \in S$. For a measure μ on (S, \mathcal{S}) and kernels $K, K' : S \times S' \rightarrow \mathbb{R}$, let $\mu K(dy) \triangleq \int \mu(dx) K(x, dy)$ and $KK'(x, dz) \triangleq \int K(x, dy) K'(y, dz)$. We will often use measures and kernels as operators. For a measurable function $\phi : S \rightarrow \mathbb{R}$, let $\mu(\phi) \triangleq \mathbb{E}_{\xi \sim \mu}[\phi(\xi)] = \int \phi(x) \mu(dx)$ and $K(x)(\phi) \triangleq \int \phi(y) K(x, dy)$. We will write $\mathbb{V}_{\xi \sim \mu}[\phi(\xi)] \triangleq \mu([\phi - \mu(\phi)]^2)$ for the variance of ϕ with respect to μ . For measures μ, ν on (S, \mathcal{S}) , we will write $\mu \ll \nu$ to denote that μ is absolutely continuous with respect to ν , in which case we will write $\frac{d\mu}{d\nu}$ for the ν -almost everywhere (ν -a.e.) unique function f satisfying $\mu(A) = \int_A f d\nu$, for all $A \in \mathcal{S}$. When the choice is clear from context, we may write $\mathcal{B}(S)$ for the σ -algebra of the space S .

Let $\mathcal{B}_b(S)$ be the set of all measurable bounded real functions on S and let $\mathcal{P}(S)$ denote the collection of all probability measures on (S, \mathcal{S}) . For $\mu, \nu \in \mathcal{P}(S)$, the total variation distance between μ and ν is

$$d_{TV}(\mu, \nu) \triangleq \sup_{A \in \mathcal{S}} |\mu(A) - \nu(A)|. \quad (1)$$

If $\mu \ll \nu$, then the KL divergence from μ to ν is

$$\text{KL}(\mu || \nu) \triangleq \mu(\log d\mu/d\nu) \quad (2)$$

and the χ^2 -divergence from μ to ν is

$$d_{\chi^2}(\mu, \nu) \triangleq \nu([d\mu/d\nu - 1]^2). \quad (3)$$

Finally, we note that, when there is little at stake, we will ignore measure-theoretic niceties such as the distinction between equality and a.e.-equality.

2.2. Problem Statement. We follow a similar setup and notation to [9]. Let $(X_t)_{t \geq 1}$ be an inhomogeneous Markov chain on (E, \mathcal{E}) with transition kernels $(M_t)_{t \geq 2}$ and initial distribution M_1 . We write $M_1(x, \cdot) = M_1(\cdot)$ when convenient. For all $t \geq 1$ and $x_{t-1} \in E$, assume that $M_t(x_{t-1}, \cdot)$ has a density with respect to some σ -finite dominating measure (which we denote by dx). We abuse notation and write the density of $M_t(x_{t-1}, \cdot)$ as $M_t(x_{t-1}, x_t)$. Denote expectations and variances with respect to the Markov chain by $\mathbb{E}[\cdot]$ and $\mathbb{V}[\cdot]$, respectively. Let $g_t : E \rightarrow \mathbb{R}_+$, for $t \geq 1$, be a sequence of \mathcal{E} -measurable potential functions on E .¹ Denote a product of potentials by $g_{s,t}(x_{s,t}) \triangleq \prod_{\tau=s}^t g_\tau(x_\tau)$ and let $g_0 \equiv 1$.

¹In the state-space setting the potential g_t would be the likelihood from the observation at time t : $g_t(x_t) = p_t(y_t | x_t)$.

The unnormalized predictive and updated measures are defined, respectively, to be

$$\gamma'_{1,t}(\phi_{1,t}) \triangleq \mathbb{E}[\phi_{1,t}(X_{1,t})g_{1,t-1}(X_{1,t-1})] = Q_{0,t}(\phi_{1,t}) \quad (4)$$

and

$$\gamma_{1,t}(\phi_{1,t}) \triangleq \mathbb{E}[\phi_{1,t}(X_{1,t})g_{1,t}(X_{1,t})] \quad (5)$$

with corresponding marginal versions

$$\gamma'_t(\phi_t) \triangleq \mathbb{E}[\phi_t(X_t)g_{1,t-1}(X_{1,t-1})] = Q_{0,t}(\phi_t) \quad (6)$$

and

$$\gamma_t(\phi_t) \triangleq \mathbb{E}[\phi_t(X_t)g_{1,t}(X_{1,t})]. \quad (7)$$

Our goal is to approximate, in a sense to be discussed later, the normalized predictive and updated measures, and their marginal versions, which are defined to be

$$\pi_{1,t}(\phi_{1,t}) \triangleq \frac{\gamma_{1,t}(\phi_{1,t})}{Z_t}, \quad \eta_{1,t}(\phi_{1,t}) \triangleq \frac{\gamma'_{1,t}(\phi_{1,t})}{Z'_t}, \quad (8)$$

$$\pi_t(\phi_t) \triangleq \frac{\gamma_t(\phi_t)}{Z_t}, \text{ and } \eta_t(\phi_t) \triangleq \frac{\gamma'_t(\phi_t)}{Z'_t}, \quad (9)$$

where $Z_t \triangleq \gamma_t(1)$ and $Z'_t \triangleq \gamma'_t(1)$ are normalization constants.

For $t \geq 1$, let $Q_t(x_{t-1}, dx_t) \triangleq g_{t-1}(x_{t-1})M_t(x_{t-1}, dx_t)$, and for $0 \leq s < t$, let

$$Q_{s,t} \triangleq Q_{s+1}Q_{s+2} \cdots Q_t, \quad (10)$$

so $Q_{t,t+1} = Q_t$. By convention $Q_{t,t}(x_t, dy_t) = \delta_{x_t}(dy_t)$ and $Q_{0,t}(dx_t)$ is a measure, not a probability kernel. Notice that for $s \in [t]$, $x_s \in E$, and $\phi_t : E \rightarrow \mathbb{R}$,

$$Q_{s,t}(x_s)(\phi_t) = \mathbb{E}[\phi_t(X_t)g_{s,t-1}(X_{s,t-1}) | X_s = x_s] \quad (11)$$

and $Q_{0,t}(\phi_t) = M_1Q_{1,t}(\phi_t)$. Generalizing these identities, we will abuse notation and write, for $s \in [t]$, $x_s \in E$, and $\phi_{s,t} : E^{t-s} \rightarrow \mathbb{R}$,

$$Q_{s,t}(x_s)(\phi_{s,t}) \triangleq \mathbb{E}[\phi_{s,t}(X_{s,t})g_{s,t-1}(X_{s,t-1}) | X_s = x_s] \quad (12)$$

and $Q_{0,t}(\phi_{1,t}) \triangleq M_1Q_{1,t}(\phi_{1,t})$. Let $G_{s,t}(y) \triangleq Q_{s,t}(y)(1)$ for $s \in [t-1]$ and $G_{0,t} \triangleq Q_{0,t}(1)$.

Observe that $\gamma'_{1,t}(\phi_{1,t}) = Q_{0,t}(\phi_{1,t})$, $\gamma'_t(\phi_t) = Q_{0,t}(\phi_t)$, and $Z'_t = Z_{t-1} = G_{0,t}$, so the one-step predictive distribution and its marginal version can be written as

$$\eta_{1,t}(\phi_{1,t}) = \frac{Q_{0,t}(\phi_{1,t})}{G_{0,t}} \quad \text{and} \quad \eta_t(\phi_t) = \frac{Q_{0,t}(\phi_t)}{G_{0,t}}. \quad (13)$$

2.3. SMC Algorithms. Sequential Monte Carlo algorithms construct approximations to the distributions $\pi_{1,t}$, π_t , $\eta_{1,t}$, and η_t , for each $t = 1, 2, \dots$ in turn, and use earlier approximations to produce later ones.

Algorithm 1 Sequential Importance Sampling

```

for  $n = 1, \dots, N$  do
  Sample  $X_1^n$  from  $M_1$ 
  Set  $X_{1,1}^n \leftarrow X_1^n$ 
end for
for  $t = 2, 3, \dots$  do
  for  $n = 1, \dots, N$  do
    Sample  $X_t^n$  from  $M_t(X_{t-1}^n, \cdot)$ 
    Set  $X_{1,t}^n \leftarrow \langle X_{1,t-1}^n, X_t^n \rangle$ 
  end for
end for
Form the SIS joint and marginal estimators  $\pi_{1,t}^{S,N}, \pi_t^{S,N}, \eta_{1,t}^{S,N}, \eta_t^{S,N}$ .

```

2.3.1. *Sequential Importance Sampling.* The SIS algorithm, given as Algorithm 1, operates by propagating a collection of N particles $\mathbf{X}_{1,t} \triangleq \{X_{1,t}^n\}_{n=1}^N$ with corresponding nonnegative weights $\mathbf{W}_t \triangleq \{W_t^n\}_{n=1}^N$, where

$$W_t^n \triangleq \frac{g_{1,t}(X_{1,t}^n)}{\sum_{k=1}^N g_{1,t}(X_{1,t}^k)}. \quad (14)$$

The updated distributions $\pi_{1,t}$ and π_t are then approximated by

$$\pi_{1,t}^{S,N} \triangleq \sum_{n=1}^N W_t^n \delta_{X_{1,t}^n} \quad \text{and} \quad \pi_t^{S,N} \triangleq \sum_{n=1}^N W_t^n \delta_{X_t^n} \quad (15)$$

and the predictive distributions $\eta_{1,t}$ and η_t are approximated by

$$\eta_{1,t}^{S,N} \triangleq \sum_{n=1}^N W_{t-1}^n \delta_{X_{1,t}^n} \quad \text{and} \quad \eta_t^{S,N} \triangleq \sum_{n=1}^N W_{t-1}^n \delta_{X_t^n}. \quad (16)$$

The estimators of the normalization constants Z_t and Z'_t are $\hat{Z}_t \triangleq \frac{1}{N} \sum_{k=1}^N g_{1,t}(X_{1,t}^k)$ and $\hat{Z}'_t \triangleq \frac{1}{N} \sum_{k=1}^N g_{1,t-1}(X_{1,t-1}^k)$. Expectations with respect the law of the SIS algorithm are written as $\mathbb{E}^{S,N}[\cdot]$.

Remark 2.1. Since all four measures of interest take very similar forms, going forward we will only explicitly define quantities related to them — such as estimators — for the marginal measures $\pi_t^{S,N}$ and $\eta_t^{S,N}$.

2.3.2. *Sampling Importance Resampling.* A more practical algorithm, which does not suffer from the weight degeneracy problem of SIS, is *sampling importance resampling* (SIR) [14, 16]. The SIR algorithm, given as Algorithm 2, is identical to SIS except for a resampling step performed after each iteration. Let $\tilde{\mathbf{W}}_t \triangleq \{\tilde{W}_t^n\}_{n=1}^N$ denote weights for the particles $\tilde{\mathbf{X}}_{1,t} \triangleq \{\tilde{X}_{1,t}^n\}_{n=1}^N$ at time t , where

$$\tilde{W}_t^n \triangleq \frac{g_t(\tilde{X}_t^n)}{\sum_{k=1}^N g_t(\tilde{X}_t^k)}.$$

The SIR marginal estimators are

$$\pi_t^{R,N} \triangleq \sum_{n=1}^N \tilde{W}_t^n \delta_{\tilde{X}_t^n} \quad \text{and} \quad \eta_t^{R,N} \triangleq \sum_{n=1}^N \frac{1}{N} \delta_{\tilde{X}_t^n} \quad (17)$$

Algorithm 2 Sampling Importance Resampling

```

for  $n = 1, \dots, N$  do
  Sample  $\tilde{X}_1^n$  from  $M_1$ 
  Set  $\tilde{X}_{1,1}^n \leftarrow \tilde{X}_1^n$ 
end for
for  $t = 2, 3, \dots$  do
  Sample  $\mathbf{A}_{t-1}$  from  $r(\cdot | \tilde{\mathbf{W}}_{t-1})$ 
  for  $n = 1, \dots, N$  do
    Sample  $\tilde{X}_t^n$  from  $M_t(\tilde{X}_{t-1}^{A_{t-1}^n}, \cdot)$ 
    Set  $\tilde{X}_{1,t}^n \leftarrow \langle \tilde{X}_{1,t-1}^{A_{t-1}^n}, \tilde{X}_t^n \rangle$ 
  end for
end for
Form the SIR joint and marginal estimators  $\pi_{1,t}^{R,N}, \pi_t^{R,N}, \eta_{1,t}^{R,N}, \eta_t^{R,N}$ .

```

and the estimators of the normalization constants Z_t and Z'_t are $\tilde{Z}_t \triangleq \frac{1}{N^t} \prod_{s=1}^t \sum_{k=1}^N g_s(X_s^k)$ and $\tilde{Z}'_t \triangleq \tilde{Z}_{t-1}$. Expectations with respect the law of the SIR algorithm are written as $\mathbb{E}^{R,N}[\cdot]$.

The SIR algorithm makes use of a resampling kernel with density $r(\mathbf{a}|\mathbf{w})$, where $\mathbf{w} \in \Delta_N \triangleq \{\mathbf{v} \in [0, 1]^N | \sum_{n=1}^N v_n = 1\}$ and $\mathbf{a} \in [N]^N$. A minimal assumption required for the algorithm to be correct is that $r(a_n = k | \mathbf{w}) = w_k$. However, we will focus on the simplest case where $r(\mathbf{a} | \mathbf{w}) = r_{\text{multi}}(\mathbf{a} | \mathbf{w}) \triangleq \prod_{n=1}^N w_{a_n}$, i.e., $a_n | \mathbf{w} \stackrel{\text{iid}}{\sim} \text{Multi}(\mathbf{w})$, for $n = 1, \dots, N$.

2.3.3. Adaptive Resampling. The SIR algorithm uses a deterministic resampling scheme. However, it is common for practitioners to use *adaptive resampling* algorithms to choose when to resample based on the realized particle weights. The most popular adaptive scheme relies on the *effective sample size* (ESS) criterion to determine when a resampling step is performed [12, 14, 20, 21]. Specifically, resampling occurs when the ESS is below some fixed threshold (e.g., $N/2$), where the ESS for unnormalized weights $\mathbf{w} = \langle w_1, \dots, w_N \rangle \in \mathbb{R}_+^N$ is

$$\mathcal{E}^2(\mathbf{w}) \triangleq \frac{(\sum_{n=1}^N w_n)^2}{\sum_{n=1}^N w_n^2} \quad (18)$$

and takes on values between 1 to N . The nomenclature arises from interpreting $\mathcal{E}^2(\mathbf{w})$ as the effective number of particles the algorithm is using if the particles have weights \mathbf{w} .

Whiteley, Lee, and Heine [23] recently introduced a very general adaptive algorithm they call α SMC, which includes SIS, SIR, and numerous other SMC variants as special cases. The α SMC algorithm, which is given as Algorithm 3, provides a flexible resampling mechanism: at each time t , a stochastic matrix α_{t-1} is chosen from a set \mathbb{A}_N of $N \times N$ matrices. We denote the value in the n -th row and k -th column of α_{t-1} by α_{t-1}^{nk} . The α SMC marginal estimators are

$$\pi_t^{\alpha,N} = \sum_{n=1}^N \frac{\ddot{W}_t^n g_t(\ddot{X}_t^n)}{\sum_{k=1}^N \ddot{W}_t^k g_t(\ddot{X}_t^k)} \delta_{\ddot{X}_t^n} \quad \text{and} \quad \eta_t^{\alpha,N} = \sum_{n=1}^N \frac{\ddot{W}_t^n}{\sum_{k=1}^N \ddot{W}_t^k} \delta_{\ddot{X}_t^n} \quad (19)$$

Algorithm 3 α SMC

```

for  $n = 1, \dots, N$  do
  Sample  $\ddot{X}_1^n$  from  $M_1$ 
  Set  $\ddot{X}_{1,1}^n \leftarrow \ddot{X}_1^n$ 
  Set  $\ddot{W}_1^n \leftarrow 1$ 
end for
for  $t = 2, 3, \dots$  do
  Select  $\alpha_{t-1}$  from  $\mathbb{A}_N$  according to some functional of  $\ddot{X}_{1,t-1}$ 
  for  $n = 1, \dots, N$  do
    Set  $\ddot{W}_t^n \leftarrow \sum_{k=1}^N \alpha_{t-1}^{nk} \ddot{W}_{t-1}^k g_{t-1}(\ddot{X}_{t-1}^k)$ 
    Sample  $A_{t-1}^n$  from  $\text{Multi}\left(\left\langle \frac{\alpha_{t-1}^{nk} \ddot{W}_{t-1}^k g_{t-1}(\ddot{X}_{t-1}^k)}{\ddot{W}_t^n} \right\rangle_{k=1}^N\right)$ 
    Sample  $\ddot{X}_t^n$  from  $M_t(\ddot{X}_{t-1}^{A_{t-1}^n}, \cdot)$ 
    Set  $\ddot{X}_{1,t}^n \leftarrow \langle \ddot{X}_{1,t-1}^n, \ddot{X}_t^n \rangle$ 
  end for
end for

```

and the estimators of the normalization constants Z_t and Z'_t are

$$\ddot{Z}_t \triangleq \frac{1}{N} \sum_{k=1}^N \ddot{W}_t^k g_t(\ddot{X}_t^k) \quad \text{and} \quad \ddot{Z}'_t \triangleq \frac{1}{N} \sum_{k=1}^N \ddot{W}_t^k. \quad (20)$$

Expectations with respect the law of the SIR algorithm are written as $\mathbb{E}^{\alpha, N}[\cdot]$.

Not only does the α SMC formulation aid in analyzing adaptive resampling strategies, it provides a useful framework for devising novel adaptive schemes with attractive computational properties, such as admitting parallelization even on resampling steps. SIS, SIR, and the standard adaptive algorithm can all be obtained as special cases of α SMC as follows. SIS is recovered when $\alpha_{t-1} = I_N$, the $N \times N$ identity matrix, while SIR is recovered when $\alpha_{t-1} = 1_{1/N}$, the $N \times N$ matrix with all entries equal to $1/N$. The so-called adaptive particle filter (APF) algorithm is obtained by setting α_{t-1} to $1_{1/N}$ if $\mathcal{E}^2(\ddot{\mathbf{W}}_{t-1}) < \zeta N$ and to I_N otherwise, where $\zeta \in (0, 1]$ is fixed.

Remark 2.2. We will write π and π^N to denote a generic updated/predictive marginal/full distribution and its associated SIS/SIR/ α SMC estimator. Expectations with respect to the estimator's law are written $\mathbb{E}^N[\cdot]$.

2.4. Conditional SMC Processes. We will need to consider variants of SIR and α SMC in which one or more particle paths is fixed ahead of time. In the case of SIR, these algorithms are known as conditional SMC algorithms [1, 2]. In the case of α SMC we will refer to the algorithms as conditional α SMC algorithms. Throughout the section, fix $t \geq 1$, $i \geq 1$, and $N \geq i$.

The *i-times conditional SMC (c^i SMC) process* is defined on the space $(E^N \times [N]^N)^{t-1} \times E^N \times [N]$, and is essentially SIR in which the first i particle trajectories $y_{1,t}^1, \dots, y_{1,t}^i \in E^t$ are a priori fixed, with lineages $\mathbf{k}^1, \dots, \mathbf{k}^i \in [N]^t$. For $\mathbf{x}_{1,t} \in$

$(E^N)^t$, $\mathbf{a}_{1,t-1} \in ([N]^N)^{t-1}$, and $a_t \in [N]$, the law of the c^i SMC process is given by

$$\mathbb{P}_{\mathbf{y}_{1,t}, \mathbf{k}_{1,t}}^{\mathbf{R}, N}(\mathbf{X}_1 \in d\mathbf{x}_1) \triangleq \prod_{j=1}^i \delta_{y_1^j}(dx_1^{k_1^j}) \prod_{\substack{n=1 \\ n \notin k_1^{1,i}}}^N M_1(dx_1^n), \quad (21)$$

for $s = 2, \dots, t$,

$$\begin{aligned} & \mathbb{P}_{\mathbf{y}_{1,t}, \mathbf{k}_{1,t}}^{\mathbf{R}, N}(\mathbf{X}_s \in d\mathbf{x}_s, \mathbf{A}_{s-1} = \mathbf{a}_{s-1} \mid \mathbf{X}_{s-1} = \mathbf{x}_{s-1}) \\ & \triangleq \prod_{j=1}^i \delta_{y_s^j}(dx_s^{k_s^j}) \mathbb{1}(a_{s-1}^{k_s^j} = k_{s-1}^j) \prod_{\substack{n=1 \\ n \notin k_s^{1,i}}}^N \frac{g_{s-1}(x_{s-1}^{a_{s-1}^n})}{\sum_{\ell=1}^N g_{s-1}(x_{s-1}^\ell)} M_s(x_{s-1}^{a_{s-1}^n}, dx_s^n). \end{aligned} \quad (22)$$

and

$$\mathbb{P}_{\mathbf{y}_{1,t}, \mathbf{k}_{1,t}}^{\mathbf{R}, N}(A_t = a_t \mid \mathbf{X}_t = \mathbf{x}_t) \triangleq \frac{g_t(x_t^{a_t})}{\sum_{n=1}^N g_t(x_t^n)}. \quad (23)$$

If $i = 1$, we write cSMC instead of c^1 SMC. Also, note that the c^i SMC process is not the same as conditioning on i trajectories and lineages of the SIR algorithm.

The *i-times conditional α SMC ($c^i\alpha$ SMC) process* is defined on the space $(E^N \times [N]^N \times [N]^i)^{t-1} \times E^N \times [N] \times [N]^i$, and is essentially α SMC in which the first i particle trajectories, but not their lineages, are a priori fixed. If $\mathbf{f} \in [N]^i$ are indices of the first i particles, let $\mathcal{D}(\mathbf{f}) \triangleq \prod_{j \neq j'} \mathbb{1}(f^j \neq f^{j'})$ be the function that indicates whether the indices are distinct. Given trajectories $y_{1,t}^1, \dots, y_{1,t}^i \in E^t$, for $\mathbf{x}_{1,t} \in (E^N)^t$, $\mathbf{f}_{1,t} \in ([N]^i)^t$, $\mathbf{a}_{1,t-1} \in ([N]^N)^{t-1}$, and $a_t \in [N]$, the law of the $c^i\alpha$ SMC process is given by

$$\mathbb{P}_{\mathbf{y}_{1,t}}^{\alpha, N}(\mathbf{X}_1 \in d\mathbf{x}_1, \mathbf{F}_1 = \mathbf{f}_1) \triangleq \mathcal{C}_1 \mathcal{D}(\mathbf{f}_1) \prod_{j=1}^i \frac{1}{N} \delta_{y_1^j}(dx_1^{f_1^j}) \prod_{n \notin f_1^{1,i}}^N M_1(dx_1^n), \quad (24)$$

for $s = 2, \dots, t$,

$$\begin{aligned} & \mathbb{P}_{\mathbf{y}_{1,t}}^{\alpha, N}(\mathbf{X}_s \in d\mathbf{x}_s, \mathbf{A}_{s-1} = \mathbf{a}_{s-1}, \mathbf{F}_s = \mathbf{f}_s \mid \\ & \quad \mathbf{X}_{1,s-1} = \mathbf{x}_{1,s-1}, \mathbf{A}_{1,s-2} = \mathbf{a}_{1,s-2}, \mathbf{F}_{s-1} = \mathbf{f}_{s-1}) \\ & \triangleq \mathcal{C}_s \mathcal{D}(\mathbf{f}_s) \prod_{j=1}^i \alpha_{s-1}^{f_s^j f_{s-1}^j} \delta_{y_s^j}(dx_s^{f_s^j}) \mathbb{1}(a_{s-1}^{f_s^j} = f_{s-1}^j) \\ & \quad \times \prod_{n \notin f_s^{1,i}} r_n(a_{s-1}^n \mid \mathbf{w}_{s-1}, \mathbf{x}_{1,s-1}) M_s(x_{s-1}^{a_{s-1}^n}, x_s^n) \end{aligned} \quad (25)$$

and

$$\mathbb{P}_{\mathbf{y}_{1,t}}^{\alpha, N}(A_t = a_t \mid \mathbf{X}_{1,t} = \mathbf{x}_{1,t}, \mathbf{A}_{1,t-1} = \mathbf{a}_{1,t-1}) \triangleq \frac{w_t^{a_t} g_t(x_t^{a_t})}{\sum_{n=1}^N w_t^n g_t(x_t^n)}, \quad (26)$$

where the \mathcal{C}_s are normalization constants that ensure the expressions are valid probabilities. We have used the notation

$$w_1^n \triangleq 1, \quad w_t^n \triangleq \sum_{k=1}^N \alpha_{t-1}^{nk} w_{t-1}^k g_{t-1}(x_{t-1}^k), \quad (27)$$

and

$$r_n(k|\mathbf{w}_{s-1}, \mathbf{x}_{1,s-1}) \triangleq \frac{\alpha_{s-1}^{nk} w_{s-1}^k g_{s-1}(x_{s-1}^k)}{w_s^n}. \quad (28)$$

If $i = 1$, we write $\text{c}\alpha\text{SMC}$ instead of $\text{c}^1\alpha\text{SMC}$.

2.5. Particle Gibbs and Iterated Conditional SMC. In the language of state-space models, the algorithms so far have been designed to approximate the posterior distribution of a Markov chain given indirect stochastic observations of the chain's values. However, it is often the case that the chain and the potentials are controlled by a global parameter $\theta \in \Theta$ for which there is a prior distribution $\varpi(d\theta)$. We replace M_s by M_s^θ and g_s by g_s^θ , then parameterize the other quantities defined previously in terms of M_s and g_s by the additional parameter θ . Throughout this section we fix t and let $(Y, \mathcal{Y}) \triangleq (E^t, \mathcal{B}(E^t))$. Since t is fixed we will suppress much of the time notation when possible in order to make the notation less cluttered. The target distribution on the product space $(\Theta \times Y, \mathcal{B}(\Theta \times Y))$ is

$$\pi(d\theta \times dy) \triangleq \gamma(d\theta \times dy)/Z, \quad (29)$$

where

$$\gamma(d\theta \times dy) \triangleq \prod_{s=1}^t g_s^\theta(y_s) M_s^\theta(y_{s-1}, dy_s) \varpi(d\theta) \quad \text{and} \quad Z \triangleq \gamma(1). \quad (30)$$

We denote the conditional distributions given $\theta \in \Theta$ or $y \in Y$ by, respectively, $\pi_\theta(dy)$ or $\pi_y(d\theta)$.

Particle Markov chain Monte Carlo (PMCMC) methods use SMC as a component within an MCMC algorithm to obtain approximate samples from $\pi(d\theta \times dy)$ [1]. We focus on the particle Gibbs (PG) sampler, which approximates the two-stage Gibbs kernel

$$\Pi(\theta, y, d\vartheta \times dz) \triangleq \pi_y(d\vartheta) \pi_\vartheta(dz). \quad (31)$$

In many settings, such as non-linear or non-Gaussian state-space models, it is possible to sample from $\pi_y(d\vartheta)$, but difficult to sample from $\pi_\vartheta(dz)$. The idea is to replace $\pi_\vartheta(dz)$ with an SMC-based approximation $\Pi_\vartheta(y, dz)$ that leaves $\pi_\vartheta(dz)$ invariant, leading to a kernel of the form $\pi_y(d\vartheta) \Pi_\vartheta(y, dz)$.

The standard PG sampler employs the iterated conditional SMC (i-cSMC) kernel [2] $P_\vartheta^{\text{R},N}$ to approximate the conditional distribution: $\Pi_\vartheta = P_\vartheta^{\text{R},N}$. For $y \in Y$, $\vartheta \in \Theta$, and with $\mathbf{1} \in \{1\}^N$, the i-cSMC kernel is given by

$$P_\vartheta^{\text{R},N}(y, dz) \triangleq \mathbb{E}_{y, \mathbf{1}, \vartheta}^{\text{R},N} \left[\delta_{X_{\mathbf{1},t}^{A_t}}(dz) \right]. \quad (32)$$

The invariant distribution of $P_\vartheta^{\text{R},N}$ is π_ϑ . The family of Markov chains with transition kernels of the form $P_\vartheta^{\text{R},N}(y, dz)$ are called i-cSMC processes.

We now introduce the α particle Gibbs (αPG) sampler, which employs what we will call the iterated conditional αSMC (i- αSMC) kernel $P_\vartheta^{\alpha,N}$ to approximate the conditional distribution: $\Pi_\vartheta = P_\vartheta^{\alpha,N}$. For $y \in Y$, $\vartheta \in \Theta$, the i- αSMC kernel, which has invariant distribution π_ϑ , is given by

$$P_\vartheta^{\alpha,N}(y, dz) \triangleq \mathbb{E}_{y, \mathbf{1}, \vartheta}^{\alpha,N} \left[\delta_{X_{\mathbf{1},t}^{A_t}}(dz) \right]. \quad (33)$$

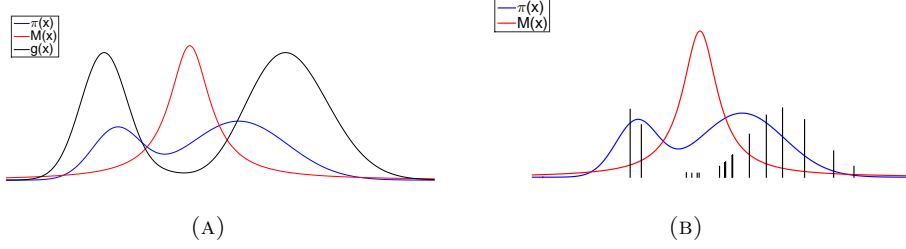


FIGURE 1. An example in which $E = \mathbb{R}$ and π and M are densities with respect to Lebesgue measure. **(a)** Plots of the densities π and M , and the potential function $g_t \propto \pi/M$. **(b)** An example of an IS estimate $\pi^{I,N}$ with $N = 20$ particles. The heights of the lines indicate the weights of the particles sampled from M .

The family of Markov chains with transition kernels of the form $P_\theta^{\alpha,N}(y, dz)$ will be called i-cSMC processes.

3. SUMMARY OF RESULTS

We now provide a summary of our results concerning SMC for sampling directly and as a component of an MCMC algorithm.

3.1. SMC for Sampling. The primary focus of our study will be on the expected value of the SMC estimators defined in Section 2.3.

Definition 3.1. If π^N is an SMC estimator of π , then the *expected estimator* is $\bar{\pi}^N \triangleq \mathbb{E}^N[\pi^N]$, where for a measurable set A , $\mathbb{E}^N[\pi^N](A) \triangleq \mathbb{E}^N[\pi^N(A)]$. \triangleleft

To connect the expected estimator to the goal of sampling we note that the marginal distribution of a sample from π^N is $\bar{\pi}^N$. To the best of our knowledge, there has been minimal investigation of expected SMC estimators, with [9, Chapter 8] a notable exception. For example, the bound

$$\text{KL}(\bar{\pi}_t^{R,N} || \pi_t) \leq \frac{c}{N}, \quad (34)$$

can be extracted as a special case of a more general propagation-of-chaos result [9, Theorem 8.3.2]. Our interest will be in the other direction of the KL divergence, $\text{KL}(\pi || \bar{\pi}^N)$. In a certain sense, the direction of the KL divergence we investigate is the more “natural,” since $\text{KL}(\mu || \nu)$ is the expected number of additional bits required to encode samples from μ when using a code for ν instead [7]. In other words, it is the amount of information lost by using samples from ν instead of samples from μ . Beyond this heuristic justification, however, we shall see that the quantities that arise when studying $\text{KL}(\pi || \bar{\pi}^N)$ are intimately related to those appearing in the study of particle Gibbs and the related iterated conditional SMC algorithm.

3.1.1. Special Case: Importance Sampling. Before detailing our results in full generality, to provide intuition as to how well $\bar{\pi}^N$ approximates π as N increases, we briefly consider the special case of importance sampling, which is equivalent to either SIS or SIR when $t = 1$. We write $M = M_1$, $g = g_1$, $\pi = \pi_1$, $Z = Z_1$, $X = X_1$,

$\pi^{I,N} = \pi_1^{S,N}$, and $\bar{\pi}^{I,N} = \bar{\pi}_1^{S,N}$. Fig. 1a gives an example of π , M , and g in the case that $E = \mathbb{R}$ and π and M are densities with respect to Lebesgue measure. Fig. 1b shows an example of an importance sampling estimate with $N = 20$ particles. Fig. 2 shows the density of $\bar{\pi}^{I,N}$, along with π and M , for $N = 4, 8, 16, 32$ particles. Informally, for a “small” number of particles, $\bar{\pi}^{I,N}$ is strongly distorted toward the proposal distribution M : for N not too large, with non-trivial probability all the samples from M will be in a region of low π -probability. Hence, all the potentials (normalized by a factor of $1/Z$) will be small ($\ll 1$). But in order to form the probability measure $\pi^{I,N}$ the sum of the weights is normalized to 1, creating an overweighting in regions of high M -probability. However, as N increases, the probability of producing a sample from M in a region of high π -probability (and thus with a large potential) increases, which induces a better approximation to π .

Using KL divergence to quantify the distortion in $\bar{\pi}^{I,N}$, we have the following:

Theorem 3.2.

$$\text{KL}(\pi || \bar{\pi}^{I,N}) \leq \log \left(1 + \frac{\mathbb{V}[Z^{-1}g(X)]}{N} \right) \leq \frac{\mathbb{V}[Z^{-1}g(X)]}{N}. \quad (35)$$

The variance $\mathbb{V}[Z^{-1}g(X)]$ is, in fact, the χ^2 -divergence from π to M since

$$\mathbb{V}[Z^{-1}g(X)] = \int (Z^{-1}g(x) - 1)^2 M(dx) \quad (36)$$

$$= \int \left(\frac{d\pi}{dM}(x) - 1 \right)^2 M(dx) \quad (37)$$

$$= d_{\chi^2}(\pi, M). \quad (38)$$

A classical result for bounding the KL divergence in terms of the χ^2 divergence can be recovered by taking $N = 1$. We then have $\text{KL}(\pi || \bar{\pi}^{I,N}) = \text{KL}(\pi || M)$, and so Theorem 3.2 implies that

$$\text{KL}(\pi || M) \leq \log(1 + d_{\chi^2}(\pi, M)). \quad (39)$$

3.1.2. Basic SMC Results. Many of our results concerning $\bar{\pi}^N$ can be seen as analogous to existing operator-perspective results, but from the measure perspective. To make the analogies transparent, we first give an operator-perspective result, followed by the comparable measure-perspective result.

A typical L^p bound for SIR states that, for any $p \geq 1$ and any $\phi \in \mathcal{B}_b(E)$ [9]

$$\mathbb{E}^{R,N} \left[|\pi_{1,t}^{R,N}(\phi) - \pi_{1,t}(\phi)|^p \right]^{1/p} \leq \frac{a(p)b(\phi)c_t}{\sqrt{N}}, \quad (40)$$

where c_t is a constant that depends only on $\{M_s, g_s\}_{s \in [t]}$. We will show that for SIS

$$\text{KL}(\pi_{1,t} || \pi_{1,t}^{S,N}) \leq \frac{\mathcal{S}_t}{N} \quad (41)$$

and for SIR

$$\text{KL}(\pi_{1,t} || \pi_{1,t}^{R,N}) \leq \frac{\mathcal{R}_t}{N} + \Theta(N^{-2}), \quad (42)$$

where \mathcal{S}_t and \mathcal{R}_t are constants depending only on $\{M_s, g_s\}_{s \in [t]}$. All these bounds hold under very mild assumptions.

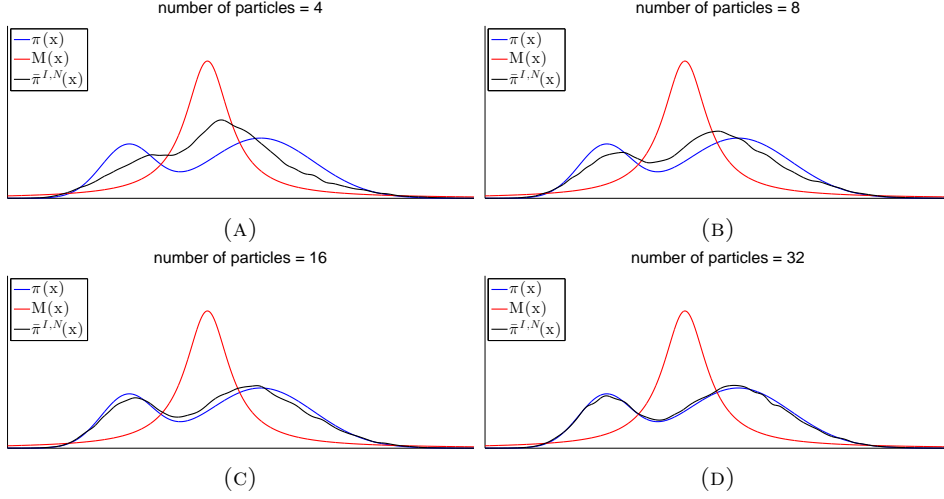


FIGURE 2. The expected IS distribution $\bar{\pi}^{I,N}$ for $N = 4, 8, 16, 32$ particles. For small numbers of particles $\bar{\pi}^{I,N}$ is strongly biased toward the proposal distribution M . The density of $\bar{\pi}^{I,N}$ in each plot was approximated using a kernel density estimate.

3.1.3. *Adaptive SMC Results.* Recall (cf. Section 2.3.3) that the effective sample size (ESS) of the α SMC particle weights at time t is

$$\mathcal{E}_t^2 \triangleq \frac{(\sum_{n=1}^N \ddot{W}_t^n)^2}{\sum_{n=1}^N (\ddot{W}_t^n)^2} \quad (43)$$

and the so-called adaptive particle filter resamples particles when $\mathcal{E}_t^2 < \zeta N$ for some fixed $\zeta \in (0, 1]$. There are myriad heuristic arguments for using the ESS criterion [20–22] as well as some theoretical analyses of the behavior of adaptive resampling algorithms under a variety of technical assumptions [8, 10, 23]. Whiteley, Lee, and Heine [23] provided a rigorous justification for the use of ESS from the operator viewpoint. Roughly speaking, they showed that if the ESS is not allowed to fall below ζN , for a fixed parameter $\zeta \in (0, 1]$, then the SMC algorithm does in fact behave as if there are ζN particles. More formally, for $\phi \in \mathcal{B}_b(E)$ and $p \geq 1$, under appropriate regularity conditions,

$$\sup_{t \geq 1} \mathcal{E}_t^2 \geq \zeta N \implies \sup_{t \geq 1} \mathbb{E}^{\alpha, N} \left[|\eta_t^{\alpha, N}(\phi) - \eta_t(\phi)|^p \right]^{1/p} \leq \frac{a(p)b(\phi)c_\infty}{\sqrt{\zeta N}}. \quad (44)$$

Comparing Eq. (44) to [9, Theorem 7.4.4], which states that

$$\sup_{t \geq 1} \mathbb{E}^{R, N} \left[|\eta_t^{R, N}(\phi) - \eta_t(\phi)|^p \right]^{1/p} \leq \frac{a(p)b(\phi)c_\infty}{\sqrt{N}}, \quad (45)$$

we see that the condition $\sup_{t \geq 1} \mathcal{E}_t^2 \geq \zeta N$ ensures that the effective number of particles in the time-uniform L^p error bound is ζN compared to N particles if SIR is used. So in this technical sense ESS is a measure of the effective sample size.

We show that a different notion of ESS leads to similar results for α SMC from the measure perspective, though we focus on fixed-time bounds, whereas the bounds

of [23] are time-uniform. We define the class of p -ESS measures, where $p \in [1, \infty]$ is the parameter of the ESS measure. For $p \in (1, \infty]$, the p -ESS is defined to be

$$\mathcal{E}_t^p \triangleq \left(\frac{\|\ddot{\mathbf{W}}_t\|_1}{\|\ddot{\mathbf{W}}_t\|_p} \right)^{p/(p-1)}, \quad (46)$$

where if $p = \infty$, then $p/(p-1) \triangleq 1$. The standard ESS given by Eq. (43) corresponds to 2-ESS. We prove that, under standard regularity conditions,

$$\sup_{s \in [t]} \mathcal{E}_s^\infty \geq \zeta N \implies \text{KL}(\pi_{1,t} \| \bar{\pi}_{1,t}^{\alpha,N}) \leq \frac{\mathcal{R}'_t}{\zeta N} + \Theta(N^{-2}), \quad (47)$$

where \mathcal{R}'_t is a constant depending only on $\{M_s, g_s\}_{s \in [t]}$.

3.2. Sampling with PG and i-cSMC. In [2], conditions are given under which the i-cSMC process is uniformly ergodic and the PG sampler is geometrically ergodic. Specifically, under regularity conditions, there exists $\rho_{t,N} = O(1/N)$ such that for all $y \in Y$ and for $P = P_\theta^{\text{R},N}$,

$$d_{TV}(\delta_y P^k, \pi_\theta) \leq \rho_{t,N}^k \quad (48)$$

and the PG sampler is geometrically ergodic as soon as the Gibbs sampler is geometrically ergodic. Furthermore, under a mixing-type condition, for any $C > 0$, if the number of particles is chosen to be $N = Ct$, then there exists $0 < \rho_C < 1$ such that $\sup_{t \geq 1} \rho_{t,N} < \rho_C$.

We give similar results for the i- α SMC process and the α PG sampler. We show that under appropriate regularity conditions, there exists $\rho_{t,N} = O(1/N)$ such that for all $y \in Y$ and for $P = P_\theta^{\alpha,N}$,

$$\sup_{s \in [t]} \mathcal{E}_s^\infty \geq \zeta N \implies \begin{cases} d_{TV}(\delta_y P^k, \pi_\theta) \leq \rho_{t,N}^k \\ \text{and} \\ \text{the } \alpha\text{PG sampler is geometrically ergodic} \\ \text{as soon as the Gibbs sampler is.} \end{cases} \quad (49)$$

Furthermore, under a mixing-type condition and a regularity condition on the matrices $\alpha \in \mathcal{A}_N$, for any $C > 0$, if $N = Ct$, then there exists $0 < \rho_C < 1$ such that $\sup_{t \geq 1} \rho_{t,N} < \rho_C$. As in sampling with α SMC, maintaining a lower bound on the ∞ -ESS is an important ingredient to the proof guaranteeing the convergence of the i- α SMC process and the α PG sampler.

4. BASIC RESULTS

In this section we give convergence rates of KL divergences for $\bar{\pi}_{1,t}^{\text{S},N}$ and $\bar{\pi}_{1,t}^{\text{R},N}$.

4.1. Convergence Rates. The key quantities in this section are

$$\begin{aligned} \mathcal{S}_t &\triangleq \mathbb{V} [Z_t^{-1} g_{1,t}(X_{1,t})], & \mathcal{R}_t &\triangleq Z_t^{-1} \sum_{s=1}^t G_{0,s} \pi_{1,t}(G_{s,t+1}) - t, \\ \mathcal{G}_t &\triangleq Z_t^{-1} \int \mathbb{E}_{y_{1,t},1}^{\text{R},N} [\tilde{Z}_t] \pi_{1,t}(dy_{1,t}). \end{aligned} \quad (50)$$

Recall that for $x_s \in E$, $G_{s,t}(x_s) = \mathbb{E}[g_{s,t-1}(X_{s,t-1}) | X_s = x_s]$. For $1 \leq \ell \leq s \leq t$, let

$$\mathcal{T}_{\ell,s} \triangleq \{\langle \tau_1, \dots, \tau_\ell \rangle : t-s+1 < \tau_1 < \dots < \tau_\ell = t+1\} \quad (51)$$

and, for $\tau \in \mathcal{T}_{\ell,s}$, define

$$C_\ell(\tau, y_{1,t}) \triangleq \prod_{i=1}^{\ell-1} G_{\tau_i, \tau_{i+1}}(y_{\tau_i}). \quad (52)$$

We will sometimes write $C_\ell^y(\tau)$ or $C_\ell^\tau(y_{1,t})$ instead of $C_\ell(\tau, y_{1,t})$. We will show (Proposition 4.8) that

$$\mathcal{G}_t = \frac{Z_t^{-1}}{N^t} \sum_{\ell=1}^{t+1} (N-1)^{t+1-\ell} \sum_{\tau \in \mathcal{T}_{\ell,t+1}} G_{0,\tau_1} \pi_{1,t}(C_\ell^\tau). \quad (53)$$

Hence, \mathcal{G}_t is related to \mathcal{R}_t :

$$\mathcal{G}_t = \frac{(N-1)^t}{N^t} + \frac{(N-1)^{t-1}}{N^t} Z_t^{-1} \sum_{s=1}^t G_{0,s} \pi_{1,t}(G_{s,t+1}) + \Theta(N^{-2}) \quad (54)$$

$$= 1 + \frac{\mathcal{R}_t}{N} + \Theta(N^{-2}). \quad (55)$$

Theorem 4.1. *For SIS,*

$$\text{KL}(\pi_{1,t} || \bar{\pi}_{1,t}^{S,N}) \leq \log \left(1 + \frac{\mathcal{S}_t}{N} \right) \leq \frac{\mathcal{S}_t}{N} \quad (56)$$

and for SIR,

$$\begin{aligned} \text{KL}(\pi_{1,t} || \bar{\pi}_{1,t}^{R,N}) &\leq \log \mathcal{G}_t = \log \left(1 + \frac{\mathcal{R}_t}{N} + \Theta(N^{-2}) \right) \\ &\leq \frac{\mathcal{R}_t}{N} + \Theta(N^{-2}). \end{aligned} \quad (57)$$

Pinsker's inequality can be used to bound the total variation distance, though the SIR convergence rate is not optimal since, as [9] shows, $d_{TV}(\pi, \bar{\pi}^{R,N}) = O(1/N)$:

Corollary 4.2.

$$d_{TV}(\pi, \bar{\pi}^{S,N}) \leq \sqrt{\frac{1}{2} \log \left(1 + \frac{\mathcal{S}_t}{N} \right)} \leq \sqrt{\frac{\mathcal{S}_t}{2N}}.$$

and

$$d_{TV}(\pi, \bar{\pi}^{R,N}) \leq \sqrt{\frac{1}{2} \log \left(1 + \frac{\mathcal{R}_t}{N} + \Theta(N^{-2}) \right)} \leq \sqrt{\frac{\mathcal{R}_t}{2N}} + \Theta(N^{-1}).$$

The following technical lemma will repeatedly prove useful:

Lemma 4.3. *Let X and Y be random elements in Borel spaces (S, \mathcal{S}) and (T, \mathcal{T}) , respectively, let $\psi : S \times T \rightarrow \mathbb{R}_+$ be a measurable, and let μ be the distribution of X . If*

$$\nu = \mathbb{E}[\psi(X, Y) \delta_X], \quad (58)$$

then $\nu \ll \mu$ and

$$\frac{d\nu}{d\mu}(X) = \mathbb{E}[\psi(X, Y) | X] \text{ a.s.} \quad (59)$$

Proof. Because S is Borel, there exists an f satisfying $f(X) = \mathbb{E}[\psi(X, Y) | X]$ a.s. It follows from the chain rule of conditional expectation and then some elementary manipulations that, for all $A \in \mathcal{S}$,

$$\nu(A) = \mathbb{E}[f(X)\delta_X(A)] = \mathbb{E}[f(X)\mathbf{1}_A(X)] = \int_A f(x)\mu(dx),$$

and so f is a version of the Radon-Nikodym derivative $d\nu/d\mu$. \square

Proposition 4.4. *For the SIS algorithm, $\bar{\pi}_{1,t}^{S,N} \ll \pi_{1,t}$ and*

$$\frac{d\bar{\pi}_{1,t}^{S,N}}{d\pi_{1,t}}(y_{1,t}) = \mathbb{E}^{S,N} \left[\frac{Z_t}{\hat{Z}_t} \middle| X_{1,t}^1 = y_{1,t} \right]. \quad (60)$$

Although Proposition 4.6 (and Proposition 4.4 below) follow as special cases of Proposition 5.1, the proofs of these results contain key ideas in simplified form, so we have included them.

Proof. We have

$$\bar{\pi}_{1,t}^{S,N} = \mathbb{E}^{S,N} \left[\sum_{n=1}^N \frac{g_{1,t}(X_{1,t}^n)}{\sum_{k=1}^N g_{1,t}(X_{1,t}^k)} \delta_{X_{1,t}^n} \right] \quad (61)$$

$$= \mathbb{E}^{S,N} \left[\frac{g_{1,t}(X_{1,t}^1)}{N^{-1} \sum_{k=1}^N g_{1,t}(X_{1,t}^k)} \delta_{X_{1,t}^1} \right] \quad (62)$$

and so, by Lemma 4.3 and the definition of \hat{Z}_t ,

$$\frac{d\bar{\pi}_{1,t}^{S,N}}{dM_{1,t}}(y_{1,t}) = g_{1,t}(y_{1,t}) \cdot \mathbb{E}^{S,N} \left[\frac{1}{\hat{Z}_t} \middle| X_{1,t}^1 = y_{1,t} \right], \quad (63)$$

where $M_{1,t} \triangleq M_1 M_2 \cdots M_t$. Upon noting that $\frac{d\bar{\pi}_{1,t}^{S,N}}{dM_{1,t}}(y_{1,t}) = g_{1,t}(y_{1,t})/Z_t$, the result follows. \square

Lemma 4.5. *$\pi_{1,t}$ and $\bar{\pi}_{1,t}^{S,N}$ are absolutely continuous with respect to each other.*

Proof. By Proposition 4.4, it suffices to show that $\pi_{1,t} \ll \bar{\pi}_{1,t}^{S,N}$. Note that for all $A \in \mathcal{B}(E^t)$, $\bar{\pi}_{1,t}^{S,N}(A) = 0 \implies$ there exists $B \subset A$ such that $M_{1,t}(B) = 0$ and $\forall y_{1,t} \in A \setminus B, g_{1,t}(y_{1,t}) = 0$. But since $\pi_{1,t} \ll M_{1,t}$, $M_{1,t}(B) = 0 \implies \pi_{1,t}(B) = 0$ and since for $y_{1,t} \in A \setminus B, g_{1,t}(y_{1,t}) = 0$, $\pi_{1,t}(A \setminus B) = 0$ as well. So $\pi_{1,t}(A) = 0$. \square

Proof of Theorem 4.1 (SIS portion). By Proposition 4.4 and Jensen's inequality

$$\frac{d\bar{\pi}_{1,t}^{S,N}}{d\pi_{1,t}}(y_{1,t}) = \mathbb{E}^{S,N} \left[\frac{Z_t}{\hat{Z}_t} \middle| X_{1,t}^1 = y_{1,t} \right] \quad (64)$$

$$\geq \frac{N}{\mathbb{E}^{S,N} \left[\sum_{k=1}^N Z_t^{-1} g_{1,t}(X_{1,t}^k) \middle| X_{1,t}^1 = y_{1,t} \right]} \quad (65)$$

$$= \frac{N}{N - 1 + Z_t^{-1} g_{1,t}(y_{1,t})}. \quad (66)$$

Lemma 4.5 implies that $\frac{d\pi_{1,t}}{d\bar{\pi}_{1,t}^{S,N}} = (\frac{d\bar{\pi}_{1,t}^{S,N}}{d\pi_{1,t}})^{-1}$, which together with Jensen's inequality yields

$$\text{KL}(\pi_{1,t} || \bar{\pi}_{1,t}^{S,N}) = \pi_{1,t} \left(\log \frac{d\pi_{1,t}}{d\bar{\pi}_{1,t}^{S,N}} \right) \leq \log \pi_{1,t} \left(\frac{d\pi_{1,t}}{d\bar{\pi}_{1,t}^{S,N}} \right) \quad (67)$$

$$= \log M_{1,t} \left(\frac{d\pi_{1,t}}{d\bar{\pi}_{1,t}^{S,N}} \frac{d\pi_{1,t}}{dM_{1,t}} \right) \quad (68)$$

$$\leq \log M_{1,t} \left(\frac{N-1 + Z_t^{-1} g_{1,t}}{N} Z_t^{-1} g_{1,t} \right) \quad (69)$$

$$= \log \left(1 + \frac{M_{1,t}(Z_t^{-1} g_{1,t})^2 - 1}{N} \right) \quad (70)$$

$$= \log \left(1 + \frac{\mathbb{V}[Z_t^{-1} g_{1,t}(X_{1,t})]}{N} \right). \quad (71)$$

□

To prove the SIR portion of Theorem 4.1 a few additional definitions are needed. We can write the joint density of SIR as

$$\psi(\mathbf{x}_{1,t}, \mathbf{a}_{1,t-1}) \triangleq \left(\prod_{n=1}^N M_1(x_1^n) \right) \prod_{s=2}^t \left(r(\mathbf{a}_{s-1} | \mathbf{w}_{s-1}) \prod_{n=1}^N M_s(x_{s-1}^{a_{s-1}^n}, x_s^n) \right), \quad (72)$$

where

$$w_s^n \triangleq w_s^n(\mathbf{x}_s) \triangleq \frac{g_s(x_s^n)}{\sum_{k=1}^n g_s(x_s^k)} \quad (73)$$

and the carrier measure is implicit. For trajectory $y_{1,t}^1 \in E^t$ and lineage $\mathbf{k}_{1,t}^1 \in [N]^t$, we can write the density of the cSMC process with law $\mathbb{P}_{y_{1,t}^1, \mathbf{k}_{1,t}^1}^{R,N}(\mathbf{X}_{1,t}, \mathbf{A}_{1,t-1})$ as

$$\begin{aligned} \tilde{\psi}_{y_{1,t}^1, \mathbf{k}_{1,t}^1}(\mathbf{x}_{1,t}, \mathbf{a}_{1,t-1}) &= \prod_{s=1}^t \mathbb{1}(a_s^1 = k_s^1) \mathbb{1}(y_s^1 = x_s^{k_s^1}) \left(\prod_{n \neq k_1^1}^N M_1(x_1^n) \right) \\ &\times \prod_{s=2}^t \left(r(\mathbf{a}_{s-1} | \mathbf{w}_{s-1}) \prod_{n \neq k_t^1}^N M_s(x_{s-1}^{a_{s-1}^n}, x_s^n) \right). \end{aligned} \quad (74)$$

Let $b_t^k \triangleq k$ and for $s = t-1, \dots, 1$, $b_s^k \triangleq a_s^{b_{s+1}^k}$, so $x_{1,t}^k = (x_1^{b_1^k}, x_2^{b_2^k}, \dots, x_t^{b_t^k})$. Furthermore, if $k = k_t^1$, then $k_s^1 = b_s^k$ and, by implicitly identifying $x_{1,t}^k$ with $y_{1,t}^1$ and $\mathbf{b}_{1,t}^k$ with $\mathbf{k}_{1,t}^1$, we can rewrite the cSMC density in the ‘‘collapsed’’ form

$$\begin{aligned} \tilde{\psi}(k, \mathbf{x}_{1,t}, \mathbf{a}_{1,t-1}) &\triangleq \left(\prod_{n \neq b_1^k}^N M_1(x_1^n) \right) \prod_{s=2}^t \left(r(\mathbf{a}_{s-1} | \mathbf{w}_{s-1}) \prod_{n \neq b_t^k}^N M_s(x_{s-1}^{a_{s-1}^n}, x_s^n) \right) \\ &= \frac{\psi(\mathbf{x}_{1,t}, \mathbf{a}_{1,t-1})}{M_1(x_1^{b_1^k}) \prod_{s=2}^t r(b_{s-1}^k | \mathbf{w}_{s-1}) M_s(x_{s-1}^{b_{s-1}^k}, x_s^{b_s^k})}. \end{aligned} \quad (75)$$

For notational convenience, we will write $\mathbb{E}_{y_{1,t}}^{\mathbf{R},N}[\cdot] \triangleq \mathbb{E}_{y_{1,t},\mathbf{1}}^{\mathbf{R},N}[\cdot]$ for expectations with respect to the law of the cSMC process with the trajectory $y_{1,t}^1 = y_{1,t}$ and the fixed lineage $\mathbf{k}_{1,t}^1 = \mathbf{1}$, the vector of all ones.

Proposition 4.6.

$$\frac{d\bar{\pi}_{1,t}^{\mathbf{R},N}}{d\pi_{1,t}}(y_{1,t}) = \mathbb{E}_{y_{1,t}}^{\mathbf{R},N} \left[\frac{Z_t}{\tilde{Z}_t} \right] \geq \frac{Z_t}{\mathbb{E}_{y_{1,t}}^{\mathbf{R},N} [\tilde{Z}_t]}. \quad (76)$$

Proof. Consider the density

$$\tilde{\pi}_{1,t}(k, \mathbf{x}_{1,t}, \mathbf{a}_{1,t-1}) \triangleq \frac{\pi_{1,t}(x_{1,t}^k) \tilde{\psi}(k, \mathbf{x}_{1,t}, \mathbf{a}_{1,t-1})}{N^t}. \quad (77)$$

Then (similarly to [1, Theorem 2])

$$\frac{\psi(\mathbf{x}_{1,t}, \mathbf{a}_{1,t-1}) w_t^k}{\tilde{\pi}_{1,t}(k, \mathbf{x}_{1,t}, \mathbf{a}_{1,t-1})} = \frac{w_t^k M_1(x_1^{b_1^k}) \prod_{s=2}^t r_{\text{multi}}(b_{s-1}^k | \mathbf{w}_{s-1}) M_s(x_{s-1}^{b_{s-1}^k}, x_s^{b_s^k})}{N^{-t} \pi_{1,t}(x_{1,t}^k)} \quad (78)$$

$$= \frac{M_1(x_1^{b_1^k}) \prod_{s=2}^t M_s(x_{s-1}^{b_{s-1}^k}, x_s^{b_s^k}) \prod_{s=1}^t w_s^{b_s^k}}{N^{-t} \pi_{1,t}(x_{1,t}^k)} \quad (79)$$

$$= \frac{M_1(x_1^{b_1^k}) \prod_{s=2}^t M_s(x_{s-1}^{b_{s-1}^k}, x_s^{b_s^k}) \prod_{s=1}^t g_s(x_{1,s}^{b_s^k})}{\pi_{1,t}(x_{1,t}^k) N^{-t} \prod_{s=1}^t \sum_{n=1}^N g_s(x_{1,s}^{b_s^n})} \quad (80)$$

$$= \frac{Z_t}{\tilde{Z}_t(\mathbf{x}_{1,t}, \mathbf{a}_{1,t-1})}, \quad (81)$$

where

$$\tilde{Z}_t(\mathbf{x}_{1,t}, \mathbf{a}_{1,t-1}) \triangleq \mathbb{E}^{\mathbf{R},N}[\tilde{Z}_t | \tilde{\mathbf{X}}_{1,t} = \mathbf{x}_{1,t}, \mathbf{A}_{1,t-1} = \mathbf{a}_{1,t-1}]. \quad (82)$$

We thus have

$$\begin{aligned} & \bar{\pi}_{1,t}^{\mathbf{R},N}(dy_{1,t}) \\ &= \sum_{\mathbf{a}_{1,t-1}, k} \int \psi(\mathbf{x}_{1,t}, \mathbf{a}_{1,t-1}) w_t^k \delta_{x_{1,t}^k}(dy_{1,t}) d\mathbf{x}_{1,t} \end{aligned} \quad (83)$$

$$= \sum_{\mathbf{a}_{1,t-1}, k} \int \frac{\psi(\mathbf{x}_{1,t}, \mathbf{a}_{1,t-1}) w_t^k}{\tilde{\pi}_{1,t}(k, \mathbf{x}_{1,t}, \mathbf{a}_{1,t-1})} \tilde{\pi}_{1,t}(k, \mathbf{x}_{1,t}, \mathbf{a}_{1,t-1}) \delta_{x_{1,t}^k}(dy_{1,t}) d\mathbf{x}_{1,t} \quad (84)$$

$$= \sum_{\mathbf{a}_{1,t-1}, k} \int \left\{ \frac{Z_t}{\tilde{Z}_t(\mathbf{x}_{1,t}, \mathbf{a}_{1,t-1})} \tilde{\pi}_{1,t}(k, \mathbf{x}_{1,t}, \mathbf{a}_{1,t-1}) \delta_{x_{1,t}^k}(dy_{1,t}) \right\} \quad (85)$$

$$= \left\{ N^{-t+1} \sum_{\mathbf{a}_{1,t-1}} \int \frac{Z_t}{\tilde{Z}_t(\mathbf{x}_{1,t}, \mathbf{a}_{1,t-1})} \tilde{\psi}(1, \mathbf{x}_{1,t}, \mathbf{a}_{1,t-1}) \delta_{y_{1,t}}(dx_{1,t}^1) \right\} \pi_{1,t}(dy_{1,t}). \quad (86)$$

The result follows from Lemma 4.3. \square

Definition 4.7. A collection of lineages $\mathbf{k}^1, \dots, \mathbf{k}^i \in [N]^t$ are said to be *distinct* when $k_s^j \neq k_s^{j'}$ for all $s \in [t]$ and $j, j' \in [i]$. \triangleleft

The next result is a straightforward generalization of [2, Proposition 9].

Proposition 4.8. *For all $t \geq 1$, $i \geq 1$, $N \geq i$, $y_{1,t}^1, \dots, y_{1,t}^i \in E^t$, and distinct lineages $\mathbf{k}_{1,t}$,*

$$\mathbb{E}_{\mathbf{y}_{1,t}, \mathbf{k}_{1,t}}^{\mathbf{R}, N}[\tilde{Z}_t] = \frac{1}{N^t} \sum_{\ell=1}^{t+1} (N-i)^{t+1-\ell} \sum_{\boldsymbol{\tau} \in \mathcal{T}_{\ell, t+1}} G_{0, \tau_1} \prod_{m=1}^{\ell-1} \sum_{j=1}^i G_{\tau_m, \tau_{m+1}}(y_{\tau_m}^j). \quad (87)$$

In particular, for $y_{1,t} \in E^t$,

$$\mathbb{E}_{y_{1,t}}^{\mathbf{R}, N}[\tilde{Z}_t] = \frac{1}{N^t} \sum_{\ell=1}^{t+1} (N-1)^{t+1-\ell} \sum_{\boldsymbol{\tau} \in \mathcal{T}_{\ell, t+1}} G_{0, \tau_1} C_{\ell}(\boldsymbol{\tau}, y_{1,t}). \quad (88)$$

Lemma 4.9. $\pi_{1,t}$ and $\bar{\pi}_{1,t}^{\mathbf{R}, N}$ are absolutely continuous with respect to each other.

Proof. The reasoning is analogous to the proof of Lemma 4.5. \square

Proof of Theorem 4.1 (SIR portion). The SIR bound follows from Propositions 4.6 and 4.8, Lemma 4.9, and Jensen's inequality. \square

Remark 4.1. As one should expect in the case of $g_s \equiv 1$ for $s \in [t]$,

$$\text{KL}(\pi_{1,t} || \bar{\pi}_{1,t}^{\mathbf{S}, N}) = \text{KL}(\pi_{1,t} || \bar{\pi}_{1,t}^{\mathbf{R}, N}) = 0,$$

and so SIS and SIR are equivalent from the measure perspective. Indeed, from the measure perspective, all that is required is a single sample from the chain $(X_t)_{t \geq 1}$. However, when $\pi_{1,t}^{\mathbf{S}, N}$ and $\pi_{1,t}^{\mathbf{R}, N}$ are used as estimators, $\pi_{1,t}^{\mathbf{S}, N}$ is superior to $\pi_{1,t}^{\mathbf{R}, N}$. Specifically, for $\phi \in \mathcal{B}_b(E^t)$, SIS produces N independent samples, so

$$\mathbb{V}^{\mathbf{S}, N}[\pi_{1,t}^{\mathbf{S}, N}(\phi)] = \frac{\mathbb{V}[\phi(X_{1,t})]}{N}.$$

For simplicity, consider a version of $\pi_{1,t}^{\mathbf{R}, N}$ obtained by first generating N samples from $\pi_{1,t}$, then applying multinomial resampling to obtain the final N samples. In this case it is easy to show that

$$\mathbb{V}^{\mathbf{R}, N}[\pi_{1,t}^{\mathbf{R}, N}(\phi)] = \frac{(2N-1)\mathbb{V}[\phi(X_{1,t})]}{N^2} \approx \frac{2\mathbb{V}^{\mathbf{S}, N}[\pi_{1,t}^{\mathbf{S}, N}(\phi)]}{N},$$

so SIS is superior to SIR from the operator perspective.

4.2. Quantitative Bounds. We can obtain explicit bounds in the SIR case using results from Andrieu, Lee, and Vihola [2] to derive bounds on the expected value of the normalization constant. To obtain bounds on the KL divergence from a bound on $\mathbb{E}_{y_{1,t}}^{\mathbf{R}, N}[\tilde{Z}_t]$ we use the following simple lemma, which follows immediately from the definition of \mathcal{G}_t and Theorem 4.1:

Lemma 4.10. *If $Z_t^{-1} \mathbb{E}_{y_{1,t}}^{\mathbf{R}, N}[\tilde{Z}_t] \leq B_N$ for some B_N that does not depend on $y_{1,t}$, then*

$$\text{KL}(\pi_{1,t} || \bar{\pi}_{1,t}^{\mathbf{R}, N}) \leq \log B_N.$$

In the case of bounded potential functions, a bound is straightforward to obtain. However, it requires the number of particles N to grow exponentially in t in order to remain constant. Let $\bar{g}_s \triangleq \sup_{x \in E} g_s(x)$.

Proposition 4.11. *Assume that $\bar{g}_s < \infty$ for all $s \in [t]$. Then for all $t \geq 1$, $i \geq 1$, $N \geq i$, $y_{1,t}^1, \dots, y_{1,t}^i \in E^t$, and distinct lineages $\mathbf{k}_{1,t}$,*

$$\mathbb{E}_{\mathbf{y}_{1,t}, \mathbf{k}_{1,t}}^{\mathbf{R}, N}[\tilde{Z}_t] \leq Z_t \left\{ 1 + [1 - (1 - i/N)^t] \left[Z_t^{-1} \prod_{s=1}^t \bar{g}_s - 1 \right] \right\}. \quad (89)$$

Proof. The proof is a straightforward generalization of that for [2, Proposition 12] with some additional bookkeeping for i (instead of 2) fixed trajectory lineages. \square

Corollary 4.12. *Assume that $\bar{g}_s < \infty$ for all $s \in [t]$. Then*

$$\begin{aligned} \text{KL}(\pi_{1,t} || \bar{\pi}_{1,t}^{\mathbf{R}, N}) &\leq \log \left\{ 1 + [1 - (1 - N^{-1})^t] \left[Z_t^{-1} \prod_{s=1}^t \bar{g}_s - 1 \right] \right\} \\ &\leq [1 - (1 - N^{-1})^t] \left[Z_t^{-1} \prod_{s=1}^t \bar{g}_s - 1 \right] \\ &\leq \frac{t(Z_t^{-1} \prod_{s=1}^t \bar{g}_s - 1)}{N}. \end{aligned}$$

Proof. Combine Lemma 4.10 and Proposition 4.11. \square

To obtain better dependence on t , stronger assumptions on $\{M_t, g_t\}$ are required.

Assumption 4.A. There exists a constant $\beta > 0$ such that for any $t, s \in \mathbb{N}$,

$$\sup_{x \in E} \frac{G_{0,t} G_{t,t+s}(x)}{G_{0,t+s}} \leq \beta.$$

\triangleleft

Proposition 4.13. *If Assumption 4.A holds, then for all $t \geq 1$, $i \geq 1$, $N \geq i$, $y_{1,t}^1, \dots, y_{1,t}^i \in E^t$, and distinct lineages $\mathbf{k}_{1,t}$,*

$$\mathbb{E}_{\mathbf{y}_{1,t}, \mathbf{k}_{1,t}}^{\mathbf{R}, N}[\tilde{Z}_t] \leq Z_t \left(1 + \frac{i(\beta - 1)}{N} \right)^t. \quad (90)$$

Proof. The proof is a simple generalization of that for [2, Proposition 14]. \square

Corollary 4.14. *If Assumption 4.A holds, then*

$$\text{KL}(\pi_{1,t} || \bar{\pi}_{1,t}^{\mathbf{R}, N}) \leq t \log \left(1 + \frac{\beta - 1}{N} \right) \leq \frac{(\beta - 1)t}{N}.$$

Proof. Combine Lemma 4.10 and Proposition 4.13. \square

Assumption 4.A is implied by a standard “strong mixing” condition which is often employed in SMC analyses (e.g., [9, 23]).

Assumption 4.B. There exists some $m \geq 1$ such that:

- (a) for some $1 \leq \gamma < \infty$, for any $t \geq 1$ and $x, x' \in E$ and $A \in \mathcal{E}$,

$$M_{t,t+m}(x, A) \leq \gamma M_{t,t+m}(x', A),$$

where $M_{t,t+m} \triangleq M_t M_{t+1} \cdots M_{t+m}$;

- (b) for some $1 \leq \delta < \infty$, the potentials satisfy

$$\sup_{x, x' \in E} \frac{g_t(x)}{g_t(x')} \leq \delta^{1/m}.$$

◁

Lemma 4.15 ([2, Lemma 17]). *If Assumption 4.B holds, then for all $t, s \geq 1$,*

$$\sup_{x, x' \in E} \frac{G_{t,t+s}(x)}{G_{t,t+s}(x')} \leq \gamma\delta,$$

so Assumption 4.A holds.

5. SMC WITH ADAPTIVE RESAMPLING

We now turn to the α SMC algorithmic framework to investigate the convergence properties of SMC with adaptive resampling. As summarized earlier, Whiteley, Lee, and Heine [23] show that maintaining a lower bound of $\mathcal{E}_t^2 \geq \zeta N$ guarantees time-uniform convergence of α SMC in the L^p sense at an $O(\frac{1}{\zeta N})$ rate. We will prove an analogous result from the measure perspective using a quantity \mathcal{E}_t^∞ we call the ∞ -ESS. We show that maintaining a lower bound of $\mathcal{E}_t^\infty \geq \zeta N$ suffices to guarantee *fixed-time* convergence of α SMC in KL divergence at an $O(\frac{1}{\zeta N})$ rate. Note that our result is weaker than that of Whiteley, Lee, and Heine [23] in the sense that theirs is time-uniform while ours is only for a fixed time. We will show that \mathcal{E}_t^∞ is a stricter notion of effective sample size than \mathcal{E}_t^2 .

We can write the joint density of the α SMC sampler as

$$\begin{aligned} & \psi^\alpha(\mathbf{x}_{1,t}, \mathbf{a}_{1,t-1}) \\ & \triangleq \left(\prod_{n=1}^N M_1(x_1^n) \right) \left(\prod_{s=2}^t \prod_{n=1}^N r_n(a_{s-1}^n | \mathbf{w}_{s-1}, \mathbf{x}_{1,s-1}) M_s(x_{s-1}^{a_{s-1}^n}, x_s^n) \right), \end{aligned} \quad (91)$$

where

$$r_n(k | \mathbf{w}_{t-1}, \mathbf{x}_{1,t-1}) \triangleq \frac{\alpha_{t-1}^{nk} w_{t-1}^k g_{t-1}(x_{t-1}^k)}{w_t^n}, \quad (92)$$

We will work under the following assumption:

Assumption 5.C. For all N , all $\alpha \in \mathbb{A}_N$ are doubly stochastic. ◁

Remark 5.1. Assumption 5.C is the same as Assumption (B⁺⁺) in [23], though there it is stated as assuming each α admits the uniform distribution as an invariant measure.

We begin by noting that, under Assumption 5.C, the density of the α SMC process with law $\mathbb{P}_{y_{1,t}^1}^{\alpha, N}(\mathbf{X}_{1,t}, \mathbf{A}_{1,t-1}, \mathbf{F}_{1,t})$ can be written in the following “collapsed” form, by implicitly identifying $x_1^{f_n}$ with $y_{1,t}^1$ (cf. Eq. (75)):

$$\begin{aligned} & \tilde{\psi}^\alpha(\mathbf{x}_{1,t}, \mathbf{a}_{1,t-1}, \mathbf{f}_{1,t}) \\ & = \frac{\psi^\alpha(\mathbf{x}_{1,t}, \mathbf{a}_{1,t-1}) \prod_{s=2}^t I_s \alpha_{s-1}^{f_s f_{s-1}}}{N M_1(x_1^{f_1}) \prod_{s=2}^t r_{f_s}(f_{s-1} | \mathbf{w}_{s-1}, \mathbf{x}_{1,s-1}) M_s(x_{s-1}^{f_{s-1}}, x_s^{f_s})} \\ & = \frac{1}{N} \prod_{n \neq f_1} M_1(x_1^n) \prod_{s=2}^t \left(I_s \alpha_{s-1}^{f_s f_{s-1}} \prod_{n \neq f_s} r_n(a_{s-1}^n | \mathbf{w}_{s-1}, \mathbf{x}_{1,s-1}) M_s(x_{s-1}^{a_{s-1}^n}, x_s^n) \right), \end{aligned} \quad (93)$$

where $I_s \triangleq \mathbb{1}(a_{s-1}^{f_s} = f_{s-1})$. Let \mathcal{F}_s be the σ -algebra generated by $\mathbf{X}_{1,s}$, $\mathbf{A}_{1,s-1}$, and $\mathbf{F}_{1,s}$, where by convention we let \mathcal{F}_0 be the trivial σ -algebra.

Proposition 5.1. *If Assumption 5.C holds, then the α SMC estimator satisfies*

$$\frac{d\bar{\pi}_{1,t}^{\alpha,N}}{d\pi_{1,t}}(y_{1,t}) = \mathbb{E}_{y_{1,t}}^{\alpha,N} \left[\frac{Z_t}{\ddot{Z}_t} \right] \geq \frac{Z_t}{\mathbb{E}_{y_{1,t}}^{\alpha,N} [\ddot{Z}_t]} \quad (94)$$

Proof. Consider the density

$$\tilde{\pi}_{1,t}^{\alpha}(\mathbf{x}_{1,t}, \mathbf{a}_{1,t-1}, \mathbf{f}_{1,t}) \triangleq \pi_{1,t}(x_{1,t}^{f_t}) \tilde{\psi}^{\alpha}(\mathbf{x}_{1,t}, \mathbf{a}_{1,t-1}, \mathbf{f}_{1,t}). \quad (95)$$

Then

$$\begin{aligned} & \frac{\psi^{\alpha}(\mathbf{x}_{1,t}, \mathbf{a}_{1,t-1}) g_t(x_t^{f_t}) w_t^{f_t}}{\tilde{\pi}_{1,t}^{\alpha}(\mathbf{x}_{1,t}, \mathbf{a}_{1,t-1}, \mathbf{f}_{1,t}) \sum_{n=1}^N g_t(x_t^n) w_t^n} \\ &= \frac{M_1(x_1^{f_1}) \prod_{s=2}^t r_{f_s}(f_{s-1} | \mathbf{w}_{s-1}, \mathbf{x}_{1,s-1}) M_s(x_{s-1}^{f_{s-1}}, x_s^{f_s}) g_t(x_t^{f_t}) w_t^{f_t}}{\pi_{1,t}(x_{1,t}^{f_t}) \prod_{s=2}^t I_s \alpha_{s-1}^{f_s f_{s-1}} N^{-1} \sum_{n=1}^N g_t(x_t^n) w_t^n} \end{aligned} \quad (96)$$

$$= \frac{M_1(x_1^{f_1}) \prod_{s=2}^t \alpha_{s-1}^{f_s f_{s-1}} w_{s-1}^{f_{s-1}} g_{s-1}(x_{s-1}^{f_{s-1}}) M_s(x_{s-1}^{f_{s-1}}, x_s^{f_s}) g_t(x_t^{f_t}) w_t^{f_t}}{\pi_{1,t}(x_{1,t}^{f_t}) \prod_{s=2}^t w_s^{f_s} \prod_{s=2}^t I_s \alpha_{s-1}^{f_s f_{s-1}} N^{-1} \sum_{n=1}^N g_t(x_t^n) w_t^n} \quad (97)$$

$$= \frac{\prod_{s=1}^t g_s(x_s^{f_s}) M_s(x_{s-1}^{f_{s-1}}, x_s^{f_s})}{\pi_{1,t}(x_{1,t}^{f_t}) N^{-1} \sum_{n=1}^N g_t(x_t^n) w_t^n \prod_{s=2}^t I_s} \quad (98)$$

$$= \frac{Z_t}{\ddot{Z}_t(\mathbf{x}_{1,t}, \mathbf{a}_{1,t-1})} \frac{1}{\prod_{s=2}^t I_s}, \quad (99)$$

where

$$\ddot{Z}_t(\mathbf{x}_{1,t}, \mathbf{a}_{1,t-1}) \triangleq \mathbb{E}^{\alpha,N} [\ddot{Z}_t | \ddot{\mathbf{X}}_{1,t} = \mathbf{x}_{1,t}, \mathbf{A}_{1,t-1} = \mathbf{a}_{1,t-1}]. \quad (100)$$

Using the convention that $0/0 = 0$, it follows that

$$\begin{aligned} & \bar{\pi}_{1,t}^{\alpha,N}(dy_{1,t}) \\ &= \sum_{\mathbf{a}_{1,t-1}, a_t} \int \left\{ \frac{\psi^{\alpha}(\mathbf{x}_{1,t}, \mathbf{a}_{1,t-1}) g_t(x_t^{a_t}) w_t^{a_t}}{\sum_{n=1}^N g_t(x_t^n) w_t^n} \delta_{x_{1,t}^{a_t}}(dy_{1,t}) \right\} \end{aligned} \quad (101)$$

$$\begin{aligned} &= \sum_{\mathbf{a}_{1,t-1}, \mathbf{f}_{1,t}} \int \left\{ \frac{\psi^{\alpha}(\mathbf{x}_{1,t}, \mathbf{a}_{1,t-1}) g_t(x_t^{f_t}) w_t^{f_t}}{\tilde{\pi}_{1,t}^{\alpha}(\mathbf{x}_{1,t}, \mathbf{a}_{1,t-1}, \mathbf{f}_{1,t}) \sum_{n=1}^N g_t(x_t^n) w_t^n} \right. \\ & \quad \left. \times \tilde{\pi}_{1,t}^{\alpha}(\mathbf{x}_{1,t}, \mathbf{a}_{1,t-1}, \mathbf{f}_{1,t}) \delta_{x_{1,t}^{f_t}}(dy_{1,t}) \right\} \end{aligned} \quad (102)$$

$$= \sum_{\mathbf{a}_{1,t-1}, \mathbf{f}_{1,t}} \int \left\{ \frac{Z_t}{\ddot{Z}_t(\mathbf{x}_{1,t}, \mathbf{a}_{1,t-1})} \tilde{\pi}_{1,t}^{\alpha}(\mathbf{x}_{1,t}, \mathbf{a}_{1,t-1}, \mathbf{f}_{1,t}) \delta_{x_{1,t}^{f_t}}(dy_{1,t}) \right\} \quad (103)$$

$$= \left\{ \sum_{\mathbf{a}_{1,t-1}, \mathbf{f}_{1,t}} \int \frac{Z_t}{\ddot{Z}_t(\mathbf{x}_{1,t}, \mathbf{a}_{1,t-1})} \tilde{\psi}^{\alpha}(k, \mathbf{x}_{1,t}, \mathbf{a}_{1,t-1}) \delta_{y_{1,t}}(dx_{1,t}^{f_t}) \right\} \pi_{1,t}(dy_{1,t}). \quad (104)$$

The result follows from Lemma 4.3. \square

With Proposition 5.1 in hand, our next task is to understand the quantity $\mathbb{E}_{y_{1,t}}^{\alpha,N} [\ddot{Z}_t]$. To do so, we will require a generalized notion of effective sample size, which includes two commonly used definitions as special cases.

Definition 5.2. For parameter $p \in [1, \infty]$, let $p_* \triangleq \frac{p}{p-1}$ be the conjugate exponent of p (so $1/p + 1/p_* = 1$). The p -**effective sample size** (p -ESS) at time t is

$$\mathcal{E}_t^p \triangleq \begin{cases} \left(\frac{\|\ddot{\mathbf{W}}_t\|_1}{\|\ddot{\mathbf{W}}_t\|_p} \right)^{p_*} & p > 1 \\ \frac{\|\ddot{\mathbf{W}}_t\|_1}{\prod_{n=1}^N (\ddot{W}_t^n)^{\ddot{W}_t^n / \|\ddot{\mathbf{W}}_t\|_1}} & p = 1. \end{cases} \quad (105)$$

◁

The following proposition catalogs some elementary properties of p -ESS (see Appendix A.1 for a proof).

Proposition 5.3. *The p -ESS has the following properties:*

(1) *The 1-ESS satisfies*

$$\mathcal{E}_t^1 = \lim_{p \downarrow 1} \mathcal{E}_t^p = \mathcal{E}_t^{\text{ent}} \triangleq e^{H(\ddot{\mathbf{W}}_t)}, \quad (106)$$

where $H(\mathbf{W}) \triangleq -\sum_n \frac{W^n}{\|\mathbf{W}\|_1} \log \frac{W^n}{\|\mathbf{W}\|_1}$.

(2) *For all $p \in [1, \infty]$, $1 \leq \mathcal{E}_t^p \leq N$. The lower bound is achieved if and only if all but one of the weights is zero. The upper bound is achieved if and only if all the weights are equal.*

(3) *For $1 \leq p < q \leq \infty$, $\mathcal{E}_t^p \geq \mathcal{E}_t^q \geq N^{-(1-q/p_*)} \mathcal{E}_t^p$, with equality if and only if \mathcal{E}_t^p and \mathcal{E}_t^q equal 1 or N .*

Part (1) of Proposition 5.3 shows that the 1-ESS corresponds to the entropic ESS, which is a common choice of ESS in applications [6]. Note that 2-ESS corresponds to the standard definition of ESS. Part (2) formalizes the sense in which p -ESS measures effective sample size: it always takes on values between 1 and N , and it takes on the value $k \in [N]$ when k particles have weight $1/k$. Part (3) shows that the larger p , the more stringent the notion of effective sample size p -ESS is. Furthermore, for $p < q$, p -ESS is a strictly stronger notion of ESS than q -ESS.

In order to prove our convergence result for α SMC, we will require a lower bound on the ∞ -ESS.

Assumption 5.D. There exists some $0 < \zeta \leq 1$ such that for all $s \in [t]$, $\mathcal{E}_s^\infty \geq \zeta N$. ◁

At a technical level, Whiteley, Lee, and Heine [23] uses the 2-ESS lower bound guarantee to bound the L_2 norm of the weights in terms of their L_1 norm. Similarly, we will use the ∞ -ESS lower bound guarantee to bound the sup-norm of the weights in terms of their L_1 norm. Specifically, under Assumption 5.D,

$$\zeta N \leq \mathcal{E}_s^\infty = \frac{\|\mathbf{W}_s\|_1}{\|\mathbf{W}_s\|_\infty} = \frac{\|\mathbf{W}_s\|_1}{\sup_n W_s^n}, \quad (107)$$

so for all $n \in [N]$ and $s \in [t]$, $W_s^n \leq \frac{\|\mathbf{W}_s\|_1}{\zeta N}$.

The expectation $\mathbb{E}_{y_{1,t}}^{\alpha,N}[\ddot{Z}_t]$ can be deconstructed in a similar manner to $\mathbb{E}_{y_{1,t}}^{\text{R},N}[\tilde{Z}_t]$. Recall that for $\boldsymbol{\tau} \in \mathcal{T}_{\ell,s}$,

$$C_\ell^y(\boldsymbol{\tau}) = \prod_{i=1}^{\ell-1} G_{\tau_i, \tau_{i+1}}^y. \quad (108)$$

Proposition 5.4. *If Assumptions 5.C and 5.D hold, then for all $t \geq 1$, $i \geq 1$, $N \geq i$, $y_{1,t}^1, \dots, y_{1,t}^i \in E^t$,*

$$\begin{aligned} & \mathbb{E}_{y_{1,t}}^{\alpha,N}[\ddot{Z}_t] \\ & \leq \frac{1}{N(\zeta N)^{t-1}} \sum_{\ell=1}^{t+1} \sum_{\tau \in \mathcal{T}_{\ell,t+1}} (\zeta N)^{t+1-\ell} \left(\frac{N-i}{\zeta N} \right)^{\mathbb{1}(\tau_1 > 1)} G_{0,\tau_1} \prod_{m=1}^{\ell-1} \sum_{j=1}^i G_{\tau_m, \tau_{m+1}}(y_{\tau_m}^j) \end{aligned} \quad (109)$$

In particular, in the case of $i = 1$, we have

$$\mathbb{E}_{y_{1,t}}^{\alpha,N}[\ddot{Z}_t] \leq \frac{1}{N(\zeta N)^{t-1}} \sum_{\ell=1}^{t+1} \sum_{\tau \in \mathcal{T}_{\ell,t+1}} (\zeta N)^{t+1-\ell} \left(\frac{N-1}{\zeta N} \right)^{\mathbb{1}(\tau_1 > 1)} G_{0,\tau_1} C_{\ell}^y(\tau) \quad (110)$$

See Appendix A.2 for a proof.

5.1. Quantitative Bounds. Proposition 5.4 leads to quantitative bounds very similar to those for SIR. Indeed, the SIR results are essentially a special case of the α SMC results, though with slightly worse constants in the α SMC case.

Theorem 5.5. *If Assumptions 5.C and 5.D hold, then for $t \geq 1$,*

$$\text{KL}(\pi_{1,t} || \bar{\pi}_{1,t}^{\alpha,N}) \leq \frac{Z_t^{-1} \sum_{s=1}^t G_{0,s} \pi_{1,t}(G_{s,t+1}) - 1}{\zeta N} + \Theta(N^{-2}). \quad (111)$$

Proof. We have

$$\mathbb{E}_{y_{1,t}}^{\alpha,N}[\ddot{Z}_t] \leq \frac{1}{N(\zeta N)^{t-1}} \sum_{\ell=1}^{t+1} \sum_{\tau \in \mathcal{T}_{\ell,t+1}} (\zeta N)^{t+1-\ell} \left(\frac{N-1}{\zeta N} \right)^{\mathbb{1}(\tau_1 > 1)} G_{0,\tau_1} C_{\ell}^y(\tau) \quad (112)$$

$$= \frac{Z_t(N-1)}{N} + \frac{G_{1,t+1}^y}{N} + \frac{1}{\zeta N} \sum_{s=2}^t G_{0,s} G_{s,t+1}^y + \Theta(N^{-2}) \quad (113)$$

$$\leq Z_t + \frac{\sum_{s=1}^t G_{0,s} G_{s,t+1}^y - Z_t}{\zeta N} + \Theta(N^{-2}). \quad (114)$$

The result then follows from Proposition 5.1. \square

It is now possible to give generalized versions of Propositions 4.11 and 4.13 and their corollaries:

Proposition 5.6. *Fix any $t \geq 1$ and assume that for $s \in [t]$, $\bar{g}_s \triangleq \sup_{x \in E} g_s(x) < \infty$. Then if Assumptions 5.C and 5.D hold, for all, $i \geq 1$, $N \geq i$, $y_{1,t}^1, \dots, y_{1,t}^i \in E^t$,*

$$\mathbb{E}_{y_{1,t}}^{\alpha,N}[\ddot{Z}_t] \leq Z_t \left\{ 1 + Z_t^{-1} \prod_{s=1}^t \bar{g}_s \left[\left(1 + \frac{1}{\zeta N} \right)^t - 1 \right] \right\} \quad (115)$$

and therefore, for $N \geq \zeta^{-1}$,

$$\text{KL}(\pi_{1,t} || \bar{\pi}_{1,t}^{\alpha,N}) \leq \log \left\{ 1 + Z_t^{-1} \prod_{s=1}^t \bar{g}_s \left[\left(1 + \frac{1}{\zeta N} \right)^t - 1 \right] \right\} \quad (116)$$

$$\leq Z_t^{-1} \prod_{s=1}^t \bar{g}_s \left[\left(1 + \frac{1}{\zeta N} \right)^t - 1 \right] \quad (117)$$

$$\leq \frac{Z_t^{-1} 2^t \prod_{s=1}^t \bar{g}_s}{\zeta N}. \quad (118)$$

Proof. We have

$$\mathbb{E}_{\mathbf{y}_{1,t}}^{\alpha,N}[\ddot{Z}_t] \leq \sum_{\ell=1}^{t+1} \sum_{\boldsymbol{\tau} \in \mathcal{T}_{\ell,t+1}} (\zeta N)^{-\ell+1} G_{0,\tau_1} C_{\ell}^y(\boldsymbol{\tau}) \quad (119)$$

$$\leq Z_t + \prod_{s=1}^t \bar{g}_s \sum_{\ell=2}^{t+1} \binom{t}{\ell-1} (\zeta N)^{-\ell+1} \quad (120)$$

$$= Z_t + \prod_{s=1}^t \bar{g}_s \sum_{\ell=1}^t \binom{t}{\ell} (\zeta N)^{-\ell} \quad (121)$$

$$= Z_t + \prod_{s=1}^t \bar{g}_s \left[\left(1 + \frac{1}{\zeta N} \right)^t - 1 \right]. \quad (122)$$

□

Proposition 5.7. *If Assumptions 4.A, 5.C and 5.D hold, then for all $t \geq 1$, $i \geq 1$, $N \geq i$, $y_{1,t}^1, \dots, y_{1,t}^i \in E^t$,*

$$\mathbb{E}_{\mathbf{y}_{1,t}}^{\alpha,N}[\ddot{Z}_t] \leq Z_t \left(1 + \frac{\beta}{\zeta N} \right)^t \quad (123)$$

and therefore

$$\text{KL}(\pi_{1,t} || \bar{\pi}_{1,t}^{\alpha,N}) \leq t \log \left(1 + \frac{\beta}{\zeta N} \right) \leq \frac{\beta t}{\zeta N}. \quad (124)$$

Proof. The proof mirrors that for [2, Proposition 14]. Observe that for $s \in [t+1]$, $G_{0,t+1} = G_{0,t+1} \frac{G_{0,t+1}}{G_{0,s}} = G_{0,s} \eta_s(G_{s,t+1})$, so we can write for $\ell \in [t]$, $\boldsymbol{\tau} \in \mathcal{T}_{\ell,t+1}$,

$$Z_t = G_{0,t+1} = G_{0,\tau_k} \prod_{i=1}^{\ell-1} \eta_{\tau_i}(G_{\tau_i,\tau_{i+1}}). \quad (125)$$

Combined with Assumption 4.A and writing $\bar{G}_{s,t} \triangleq \sup_{x \in E} G_{s,t}(x)$,

$$\sum_{\ell=1}^{t+1} (\zeta N)^{-\ell+1} \sum_{\boldsymbol{\tau} \in \mathcal{T}_{\ell,t+1}} G_{0,\tau_1} \prod_{i=1}^{\ell-1} G_{\tau_i,\tau_{i+1}}^y \quad (126)$$

$$\leq Z_t + Z_t \sum_{\ell=2}^{t+1} (\zeta N)^{-\ell+1} \sum_{\boldsymbol{\tau} \in \mathcal{T}_{\ell,t+1}} \frac{G_{0,\tau_1}}{G_{0,\tau_1}} \prod_{i=1}^{\ell-1} \frac{\bar{G}_{\tau_i,\tau_{i+1}}}{\eta_{\tau_i}(G_{\tau_i,\tau_{i+1}})} \quad (127)$$

$$= Z_t \sum_{\ell=1}^{t+1} \binom{t}{\ell-1} (\zeta N)^{-\ell+1} \beta^{\ell-1} \quad (128)$$

$$= Z_t \left(1 + \frac{\beta}{\zeta N} \right)^t. \quad (129)$$

□

6. CONVERGENCE IN THE MARGINAL AND PREDICTIVE CASES

So far we have focused on the convergence of $\bar{\pi}_{1,t}^{S,N}$, $\bar{\pi}_{1,t}^{R,N}$, and $\bar{\pi}_{1,t}^{\alpha,N}$, which are all expectations of estimators of $\pi_{1,t}$. Very similar results can be derived for

the expected estimators of $\pi_t, \eta_{1,t}$, and η_t . The following result allows us to easily generalize all the quantitative bounds given so far to these marginal and/or predictive measures. Define the reverse probability kernels $\tilde{\pi}_{1,t-1}(x_t, dx_{1,t-1})$ and $\tilde{\eta}_{1,t-1}(x_t, dx_{1,t-1})$ such that

$$\pi_{1,t}(dx_{1,t}) = \pi_t(dx_t) \tilde{\pi}_{1,t-1}(x_t, dx_{1,t-1}) \quad (130)$$

and

$$\eta_{1,t}(dx_{1,t}) = \eta_t(dx_t) \tilde{\eta}_{1,t-1}(x_t, dx_{1,t-1}). \quad (131)$$

Proposition 6.1. *If Assumption 5.C holds, then the expected α SMC estimators satisfy*

$$\frac{d\bar{\pi}_t^{\alpha,N}}{d\pi_t}(y_t) = \int \mathbb{E}_{y_{1,t}}^{\alpha,N} \left[\frac{Z_t}{\ddot{Z}_t} \right] \tilde{\pi}_{1,t-1}(y_t, dy_{1,t-1}) \quad (132)$$

$$\frac{d\bar{\eta}_{1,t}^{\alpha,N}}{d\eta_{1,t}}(y_{1,t}) = \mathbb{E}_{y_{1,t}}^{\alpha,N} \left[\frac{Z'_t}{\ddot{Z}'_t} \right] = \mathbb{E}_{y_{1,t}}^{\alpha,N} \left[\frac{Z_{t-1}}{\ddot{Z}_{t-1}} \right] \quad (133)$$

$$\begin{aligned} \frac{d\bar{\eta}_t^{\alpha,N}}{d\eta_t}(y_t) &= \int \mathbb{E}_{y_{1,t}}^{\alpha,N} \left[\frac{Z'_t}{\ddot{Z}'_t} \right] \tilde{\eta}_{1,t-1}(y_t, dy_{1,t-1}) \\ &= \int \mathbb{E}_{y_{1,t}}^{\alpha,N} \left[\frac{Z_{t-1}}{\ddot{Z}_{t-1}} \right] \tilde{\eta}_{1,t-1}(y_t, dy_{1,t-1}). \end{aligned} \quad (134)$$

Proof. For Eq. (132), by Proposition 5.1, we have

$$\bar{\pi}_t^{\alpha,N}(dy_t) = \int \frac{d\bar{\pi}_{1,t}^{\alpha,N}}{d\pi_{1,t}}(y_{1,t}) \pi_{1,t}(dy_{1,t}) = \int \mathbb{E}_{y_{1,t}}^{\alpha,N} \left[\frac{Z_t}{\ddot{Z}_t} \right] \pi_{1,t}(dy_{1,t}) \quad (135)$$

$$= \int \mathbb{E}_{y_{1,t}}^{\alpha,N} \left[\frac{Z_t}{\ddot{Z}_t} \right] \tilde{\pi}_{1,t-1}(y_t, dy_{1,t-1}) \pi_t(dy_t). \quad (136)$$

For Eq. (133), consider the density

$$\tilde{\eta}_{1,t}^\alpha(\mathbf{x}_{1,t}, \mathbf{a}_{1,t-1}, \mathbf{f}_{1,t}) \triangleq \eta_{1,t}(x_{1,t}^{f_t}) \tilde{\psi}^\alpha(\mathbf{x}_{1,t}, \mathbf{a}_{1,t-1}, \mathbf{f}_{1,t}). \quad (137)$$

Then

$$\begin{aligned} & \frac{\psi^\alpha(\mathbf{x}_{1,t}, \mathbf{a}_{1,t-1}) w_t^{f_t}}{\tilde{\eta}_{1,t}^\alpha(\mathbf{x}_{1,t}, \mathbf{a}_{1,t-1}, \mathbf{f}_{1,t}) \sum_{n=1}^N w_t^n} \\ &= \frac{M_1(x_1^{f_1}) \prod_{s=2}^t r_{f_s}(f_{s-1} | \mathbf{w}_{s-1}, \mathbf{x}_{1,s-1}) M_s(x_{s-1}^{f_{s-1}}, x_s^{f_s}) w_t^{f_t}}{\eta_{1,t}(x_{1,t}^{f_t}) \prod_{s=2}^t I_s \alpha_{s-1}^{f_s f_{s-1}} N^{-1} \sum_{n=1}^N w_t^n} \end{aligned} \quad (138)$$

$$= \frac{M_1(x_1^{f_1}) \prod_{s=2}^t \alpha_{s-1}^{f_s f_{s-1}} w_{s-1}^{f_{s-1}} g_{s-1}(x_{s-1}^{f_{s-1}}) M_s(x_{s-1}^{f_{s-1}}, x_s^{f_s}) w_t^{f_t}}{\eta_{1,t}(x_{1,t}^{f_t}) \prod_{s=2}^t w_s^{f_s} \prod_{s=2}^t I_s \alpha_{s-1}^{f_s f_{s-1}} N^{-1} \sum_{n=1}^N w_t^n} \quad (139)$$

$$= \frac{M_1(x_1^{f_1}) \prod_{s=2}^t g_{s-1}(x_{s-1}^{f_{s-1}}) M_s(x_{s-1}^{f_{s-1}}, x_s^{f_s})}{\eta_{1,t}(x_{1,t}^{f_t}) N^{-1} \sum_{n=1}^N w_t^n \prod_{s=2}^t I_s} \quad (140)$$

$$= \frac{Z'_t}{\ddot{Z}'_t(\mathbf{x}_{1,t}, \mathbf{a}_{1,t-1})} \frac{1}{\prod_{s=2}^t I_s}, \quad (141)$$

where

$$\ddot{Z}'_t(\mathbf{x}_{1,t}, \mathbf{a}_{1,t-1}) \triangleq \mathbb{E}^{\alpha,N}[\ddot{Z}'_t | \mathbf{X}_{1,t} = \mathbf{x}_{1,t}, \mathbf{A}_{1,t-1} = \mathbf{a}_{1,t-1}]. \quad (142)$$

We thus have

$$\begin{aligned} & \bar{\eta}_{1,t}^{\alpha,N}(\mathrm{d}y_{1,t}) \\ &= \sum_{\mathbf{a}_{1,t-1}, \mathbf{a}_t} \int \left\{ \frac{\psi^\alpha(\mathbf{x}_{1,t}, \mathbf{a}_{1,t-1}) w_t^{a_t}}{\sum_{n=1}^N w_t^n} \delta_{x_{1,t}^{a_t}}(\mathrm{d}y_{1,t}) \right\} \end{aligned} \quad (143)$$

$$\begin{aligned} &= \sum_{\mathbf{a}_{1,t-1}, \mathbf{f}_{1,t}} \int \left\{ \frac{\psi^\alpha(\mathbf{x}_{1,t}, \mathbf{a}_{1,t-1}) w_t^{f_t}}{\tilde{\eta}_{1,t}^\alpha(\mathbf{x}_{1,t}, \mathbf{a}_{1,t-1}, \mathbf{f}_{1,t}) \sum_{n=1}^N w_t^n} \right. \\ &\quad \left. \times \tilde{\eta}_{1,t}^\alpha(\mathbf{x}_{1,t}, \mathbf{a}_{1,t-1}, \mathbf{f}_{1,t}) \delta_{x_{1,t}^{f_t}}(\mathrm{d}y_{1,t}) \right\} \end{aligned} \quad (144)$$

$$= \sum_{\mathbf{a}_{1,t-1}, \mathbf{f}_{1,t}} \int \left\{ \frac{Z'_t}{\ddot{Z}'_t(\mathbf{x}_{1,t}, \mathbf{a}_{1,t-1})} \tilde{\eta}_{1,t}^\alpha(\mathbf{x}_{1,t}, \mathbf{a}_{1,t-1}, \mathbf{f}_{1,t}) \delta_{x_{1,t}^{f_t}}(\mathrm{d}y_{1,t}) \right\} \quad (145)$$

$$= \left\{ \sum_{\mathbf{a}_{1,t-1}, \mathbf{f}_{1,t}} \int \frac{Z'_t}{\ddot{Z}'_t(\mathbf{x}_{1,t}, \mathbf{a}_{1,t-1})} \tilde{\psi}^\alpha(k, \mathbf{x}_{1,t}, \mathbf{a}_{1,t-1}) \delta_{y_{1,t}}(\mathrm{d}x_{1,t}^{f_t}) \right\} \pi_{1,t}(\mathrm{d}y_{1,t}), \quad (146)$$

proving the first equality. The second equality follows since the estimator of the normalization constant can be written as

$$\ddot{Z}'_t = N^{-1} \sum_n W_t^n = N^{-1} \sum_n \sum_k \alpha_{t-1}^{nk} W_{t-1}^k g_{t-1}^k \quad (147)$$

$$= N^{-1} \sum_k W_{t-1}^k g_{t-1}^k = \ddot{Z}_{t-1}, \quad (148)$$

where the penultimate equality follows from Assumption 5.C.

Eq. (134) follows by the same argument used to prove Eq. (132), but using Eq. (133). \square

For π_t , we have bounds identical to those for $\pi_{1,t}$:

Theorem 6.2. *For the SIS and SIR algorithms,*

$$\mathrm{KL}(\pi_t || \bar{\pi}_t^{\mathrm{S},N}) \leq \log \left(1 + \frac{\mathcal{S}_t}{N} \right) \leq \frac{\mathcal{S}_t}{N} \quad (149)$$

and

$$\mathrm{KL}(\pi_t || \bar{\pi}_t^{\mathrm{R},N}) \leq \log \mathcal{G}_t. \quad (150)$$

Proof. For SIS,

$$\begin{aligned} \frac{\mathrm{d}\bar{\pi}_t^{\mathrm{S},N}}{\mathrm{d}\pi_t}(y_{1,t}) &= \int \mathbb{E}^{\mathrm{S},N} \left[\frac{N}{\sum_{n=1}^N Z_t^{-1} g_{1,t}(X_{1,t}^n)} \middle| X_{1,t}^1 = y_{1,t} \right] \tilde{\pi}_{1,t-1}(y_t, \mathrm{d}y_{1,t-1}) \\ &\geq \frac{N}{N-1 + \int Z_t^{-1} g_{1,t}(y_{1,t}) \tilde{\pi}_{1,t-1}(y_t, \mathrm{d}y_{1,t-1})}. \end{aligned} \quad (151)$$

$$\geq \frac{N}{N-1 + \int Z_t^{-1} g_{1,t}(y_{1,t}) \tilde{\pi}_{1,t-1}(y_t, \mathrm{d}y_{1,t-1})}. \quad (152)$$

Hence,

$$\begin{aligned}
\text{KL}(\pi_t || \bar{\pi}_t^{S,N}) &\leq \int \log \left(\frac{N-1 + \int Z_t^{-1} g_{1,t}(y_{1,t}) \tilde{\pi}_{1,t-1}(y_t, dy_{1,t-1})}{N} \right) \pi_t(dy_t) \\
&\leq \log \left(\int \frac{N-1 + \int Z_t^{-1} g_{1,t}(y_{1,t}) \tilde{\pi}_{1,t-1}(y_t, dy_{1,t-1})}{N} \pi_t(dy_t) \right) \\
&= \log \left(1 + \frac{\pi_{1,t}(Z_t^{-1} g_{1,t}) - 1}{N} \right) \\
&= \log \left(1 + \frac{\mathcal{S}_t}{N} \right).
\end{aligned}$$

For SIR,

$$\frac{d\bar{\pi}_t^{R,N}}{d\pi_t}(y_t) \geq \frac{Z_t}{\int \mathbb{E}_{y_{1,t}}^{R,N} [\tilde{Z}_t] \tilde{\pi}_{1,t-1}(y_t, dy_{1,t-1})}. \quad (153)$$

and the bound follows analogously to the SIS case. \square

Using Proposition 5.1 and Theorem 6.2 together with the results from Sections 4.2 and 5.1, the following quantitative bounds follow immediately:

Theorem 6.3. *For all $t \geq 1$,*

(1) if for $s \in [t]$, $\bar{g}_s < \infty$, then

$$\text{KL}(\pi_t || \bar{\pi}_t^{R,N}) \leq \frac{t(Z_t^{-1} \prod_{s=1}^t \bar{g}_s - 1)}{N}, \quad (154)$$

while if Assumptions 5.C and 5.D also hold, then for $N \geq \zeta^{-1}$

$$\text{KL}(\pi_t || \bar{\pi}_t^{\alpha,N}) \leq \frac{Z_t^{-1} 2^t \prod_{s=1}^t \bar{g}_s}{\zeta N}; \quad (155)$$

and

(2) if Assumption 4.A holds, then

$$\text{KL}(\pi_t || \bar{\pi}_t^{R,N}) \leq \frac{(\beta - 1)t}{N}, \quad (156)$$

while if Assumptions 5.C and 5.D also hold, then

$$\text{KL}(\pi_t || \bar{\pi}_t^{\alpha,N}) \leq \frac{\beta t}{\zeta N}. \quad (157)$$

Remark 6.1. From the operator perspective, $\pi_t^{S,N}$, $\pi_t^{R,N}$, and $\pi_t^{\alpha,N}$ generally approximate π_t far better than $\pi_{1,t}^{S,N}$, $\pi_{1,t}^{R,N}$, and $\pi_{1,t}^{\alpha,N}$ approximate $\pi_{1,t}$. It is quite natural for SMC to produce better estimates of the marginal expectation since, while both the marginal and joint estimators involve the same number of particles, the joint expectation involves an integral over a much higher-dimensional space. So it is somewhat surprising that the KL divergence bounds we obtain in the marginal case are identical to those in the full joint distribution case already considered. But in fact, intuition suggests that the KL divergence case will behave very differently from that of functional approximation. Since only a single sample is being drawn from $\pi_{1,t}^{S,N}$, $\pi_{1,t}^{R,N}$, or $\pi_{1,t}^{\alpha,N}$, the quality of the full sample compared to the marginal sample does not suffer from the same “curse of dimensionality.”

Using Proposition 5.1 together with the results of Sections 4.2 and 5.1, we obtain similar results for estimators of the predictive measure $\eta_{1,t}$ to those for the estimators of $\pi_{1,t}$:

Theorem 6.4. *For all $t \geq 1$,*

(1) if for $s \in [t-1]$, $\bar{g}_s < \infty$, then

$$\text{KL}(\eta_{1,t} || \bar{\eta}_{1,t}^{\text{R},N}) \leq \frac{(t-1)(Z_t'^{-1} \prod_{s=1}^{t-1} \bar{g}_s - 1)}{N}, \quad (158)$$

while if Assumptions 5.C and 5.D also hold, then for $N \geq \zeta^{-1}$

$$\text{KL}(\eta_{1,t} || \bar{\eta}_{1,t}^{\alpha,N}) \leq \frac{Z_t'^{-1} 2^t \prod_{s=1}^{t-1} \bar{g}_s}{\zeta N}; \quad (159)$$

and

(2) if Assumption 4.A holds, then

$$\text{KL}(\eta_{1,t} || \bar{\eta}_{1,t}^{\text{R},N}) \leq \frac{(\beta-1)(t-1)}{N}, \quad (160)$$

while if Assumptions 5.C and 5.D also hold, then

$$\text{KL}(\eta_{1,t} || \bar{\eta}_{1,t}^{\alpha,N}) \leq \frac{\beta(t-1)}{\zeta N}. \quad (161)$$

7. PARTICLE MCMC

We now turn to particle MCMC methods, specifically i-cSMC and particle Gibbs. We will leverage results from Section 5 to prove mixing and convergence results for adaptive versions of i-cSMC and particle Gibbs under the condition that a lower bound on the ∞ -ESS is maintained by the algorithm. Briefly, let us recall the setting and some key definitions from Section 2.5. We parameterize the Markov chain $(X_t)_{t \geq 1}$ and the potentials by a global parameter $\theta \in \Theta$ for which there is a prior distribution $\varpi(d\theta)$. Hence, M_s is replaced by M_s^θ and g_s is replaced by g_s^θ . Throughout this section we fix t and let (Y, \mathcal{Y}) be the measurable space with $Y \triangleq E^t$ and $\mathcal{Y} \triangleq \mathcal{B}(Y)$. Since t is fixed we will suppress most of the time notation. The target distribution is

$$\pi(d\theta \times dy) \triangleq \gamma(d\theta \times dy)/Z, \quad (162)$$

where

$$\gamma(d\theta \times dy) \triangleq \prod_{s=1}^t g_s^\theta(y_s) M_s^\theta(y_{s-1}, dy_s) \varpi(d\theta) \quad \text{and} \quad Z \triangleq \gamma(1). \quad (163)$$

We denote the conditional distributions given θ or y by, respectively, $\pi_\theta(dy)$ or $\pi_y(d\theta)$.

The particle Gibbs (PG) sampler uses the MCMC kernel

$$\Pi_G(\theta, y, d\vartheta \times dz) \triangleq \pi_y(d\vartheta) \Pi_\vartheta(y, dz). \quad (164)$$

First, in Section 7.1, we discuss the standard PG algorithm, where $\Pi_\vartheta = P_\vartheta^{\text{R},N}$, the i-cSMC kernel. In particular, we summarize results from [2] and describe the close technical connections between their techniques and those from Section 4. In Section 7.2 we use results from Section 5 to prove convergence and mixing results for iterated conditional α SMC and α PG, where $\Pi_\vartheta = P_\vartheta^{\alpha,N}$, the i- α SMC kernel.

7.1. Convergence of PG and i-cSMC. Recall from Section 2.5 that the i-cSMC kernel is given by

$$P_\theta^{\text{R},N}(y, dz) \triangleq \mathbb{E}_{y,1,\theta}^{\text{R},N} \left[\delta_{X_{1,t}^{A_t}}(dz) \right]. \quad (165)$$

We will first state some properties of i-cSMC processes, then the particle Gibbs sampler. For notational convenience, we will write $\mathbb{E}_{y,z,\mathbf{k},\theta}^{\text{R},N}[\cdot] \triangleq \mathbb{E}_{\mathbf{y}_{1,t},\mathbf{k}_{1,t},\theta}^{\text{R},N}[\cdot]$ for expectations with respect to the law of the c²SMC process with trajectories $y^1 = y$ and $y^2 = z$, and lineages $\mathbf{k}_{1,t}^1 = \mathbf{1}$ and $\mathbf{k}_{1,t}^2 = \mathbf{k}$.

Proposition 7.1 ([2, Proposition 6]). *For $y \in Y$ and $N \geq 2$,*

$$P_\theta^{\text{R},N}(y, dz) \geq \frac{Z(1-1/N)^t}{\mathbb{E}_{y,z,\mathbf{k},\theta}^{\text{R},N}[\tilde{Z}]} \pi_\theta(dz). \quad (166)$$

From Proposition 7.1 it follows that if $\sup_{y,z \in Y} \mathbb{E}_{y,z,\mathbf{k},\theta}^{\text{R},N}[\tilde{Z}] \leq B_N$, the cSMC kernel satisfies the minorization condition

$$P_\theta^{\text{R},N}(y, dz) \geq \varepsilon_{t,N} \pi_\theta(dz), \quad (167)$$

where $\varepsilon_{t,N} \triangleq \frac{Z(1-1/N)^t}{B_N}$. The minorization condition, in turn, implies uniform ergodicity and a number of other types of convergence guarantees for the i-cSMC process. For a stationary Markov chain $(\xi_k)_{k \geq 0}$ with μ -reversible Markov kernel K , the asymptotic variance for the function $\phi \in L^2(S, \mu)$ is defined to be

$$\mathbb{V}[\phi, K] \triangleq \lim_{k \rightarrow \infty} \mathbb{V} \left[k^{-1/2} \sum_{i=1}^k [\phi(\xi_i) - \pi(\phi)] \right]. \quad (168)$$

Proposition 7.2 ([2, Proposition 7]). *Let μ be a probability measure on (S, \mathcal{S}) and let $\Pi : S \times S \rightarrow [0, 1]$ be a probability kernel that is reversible with respect to μ . If the stationary Markov chain defined by Π is ψ -irreducible and aperiodic and there exists $\varepsilon > 0$ such that for all $y \in Y$, $\Pi(y, dz) \geq \varepsilon \mu(dz)$, then*

(1) *for any probability measure $\nu \ll \mu$ and $k \geq 1$,*

$$d_{\chi^2}(\nu \Pi^k, \mu) \leq d_{\chi^2}(\nu, \mu)(1 - \varepsilon)^k, \quad (169)$$

(2) *for any $y \in Y$ and $k \geq 1$,*

$$d_{TV}(\delta_y \Pi^k, \mu) \leq (1 - \varepsilon)^k, \quad (170)$$

and

(3) *for any $\phi \in L^2(S, \mu)$*

$$\mathbb{V}[\phi, \Pi] \leq (2\varepsilon^{-1} - 1) \mathbb{V}_{\xi \sim \mu}[\phi(\xi)]. \quad (171)$$

Corollary 7.3. *If $\sup g_s^\theta < \infty$ for all $s \in [t]$, then all the results of Proposition 7.2 apply with $\Pi = P_\theta^{\text{R},N}$ and $\varepsilon = \varepsilon_{t,N}$, where $1 - \varepsilon_{t,N} = O(1/N)$. If in addition Assumption 4.A holds, then for every $C > 0$ there exists an $\varepsilon_C > 0$ such that for all $t > 1$, if $N = Ct$, then $\varepsilon_{t,N} \geq \varepsilon_C > 0$.*

Proof. The first part follows from Propositions 7.1, 7.2 and 4.11. The second then follows from Proposition 4.13. \square

Remark 7.1. The second part of the corollary states that, if $\sup g_s^\theta < \infty$ and Assumption 4.A hold, then scaling N linearly with t ensures a uniform convergence rate, as measured by χ^2 -divergence, total variation distance, or asymptotic variance.

Recall that $\gamma_\theta(dy) = \prod_{s=1}^t g_s^\theta(y_s) M_s^\theta(y_{s-1}, dy_s)$. The results of [2, Section 7] show how to use the convergence guarantees of the i-cSMC kernel to give conditions for the geometric ergodicity of the PG sampler:

Theorem 7.4. *If*

$$\pi\text{-ess sup}_\theta \frac{\prod_{s=1}^t \bar{g}_{\theta,s}}{\gamma_\theta(1)} < \infty \quad (172)$$

or there exists $1 \leq \beta < \infty$ such that for any $t, s \in \mathbb{N}$,

$$\pi\text{-ess sup}_{\theta,x} \frac{G_{0,t} G_{t,t+s}(x)}{G_{0,t+s}} \leq \beta. \quad (173)$$

then as soon as the Gibbs sampler is geometrically ergodic, the PG chain is geometrically ergodic.

7.2. Convergence of α PG and i- α SMC. We now generalize the results of the previous section to allow for adaptive resampling. Of particular note is that our results show that ∞ -ESS is a measure of effective sample size for adaptive resampling in this setting. Recall from Section 2.5 that the i- α SMC kernel can be used to define the α PG algorithm. For $y \in Y$, i- α SMC kernel is given by

$$P_\theta^{\alpha,N}(y, dz) \triangleq \mathbb{E}_{y,1,\theta}^{\alpha,N} [\delta_{X_{1,t}^{A_t}}(dz)]. \quad (174)$$

We call the family of Markov chains with transition kernels of the form $P_\theta^{\alpha,N}(y, dz)$ i- α SMC processes.

Our primary technical goal will be to prove a minorization condition for $P_\theta^{\alpha,N}(y, dz)$ by generalizing Proposition 7.1. We begin by noting that, under Assumption 5.C, the normalization constants for the $c^2\alpha$ SMC process are given by

$$\mathcal{C}_1 \triangleq \frac{N}{N-1} \quad (175)$$

and, for $s = 2, \dots, t$,

$$\mathcal{C}_s \triangleq \left(1 - \sum_{k=1}^N \alpha_{s-1}^{kf_s^y-1} \alpha_{s-1}^{kf_s^z-1} \right)^{-1}. \quad (176)$$

Let

$$\kappa_N \triangleq \max_{n \neq n', \alpha \in \mathbb{A}_N} \sum_{k=1}^N \alpha^{kn} \alpha^{kn'} \quad \text{and} \quad \kappa'_N \triangleq \kappa_N \vee 1/N, \quad (177)$$

so $\mathcal{C}_s \leq \frac{1}{1-\kappa_N}$ for $s = 2, \dots, t$. Thus, for all $s \in [t]$, $\mathcal{C}_s \leq \frac{1}{1-\kappa'_N}$. Here $a \vee b$ denotes the maximum of $a, b \in \mathbb{R}$.

Proposition 7.5. *For $y \in Y$ and $N \geq 2$,*

$$P_\theta^{\alpha,N}(y, dz) \geq \frac{Z_t}{\mathbb{E}_{y,z,\theta}^{\alpha,N}[\ddot{Z}_t \prod_{s=1}^t \mathcal{C}_s]} \pi(dz) \geq \frac{Z_t(1-\kappa'_N)^t}{\mathbb{E}_{y,z,\theta}^{\alpha,N}[\ddot{Z}_t]} \pi(dz). \quad (178)$$

The proof of Proposition 7.5 can be found in Appendix A.3. We recover Proposition 7.1 as a special case of Proposition 7.5 since, for SIR, $\kappa'_N = 1/N$.

Using Proposition 5.4, we obtain:

Proposition 7.6. *If Assumptions 5.C and 5.D hold, then for all $N \geq 2$,*

$$\mathbb{E}_{y,z,\theta}^{\alpha,N}[\ddot{Z}_t] \leq \frac{1}{N(\zeta N)^{t-1}} \sum_{\ell=1}^{t+1} \sum_{\tau \in \mathcal{T}_{\ell,t+1}} (\zeta N)^{t-\ell} \left(\frac{N-2}{\zeta N} \right)^{\mathbf{1}(\tau_1 > 1)} G_{0,\tau_1} C_\ell^{y+z}(\tau), \quad (179)$$

where

$$C_\ell^{y+z}(\tau) \triangleq \prod_{i=1}^{\ell-1} (G_{\tau_i, \tau_{i+1}}(y_{\tau_i}) + G_{\tau_i, \tau_{i+1}}(z_{\tau_i})). \quad (180)$$

Hence, if Assumptions 5.C and 5.D hold and the potentials are bounded, then $\mathbb{E}_{y,z,\theta}^{\alpha,N}[\ddot{Z}_t] = 1 + O(N^{-1})$, while if Assumptions 4.A, 5.C and 5.D hold, then

$$\mathbb{E}_{y,z,\theta}^{\alpha,N}[\ddot{Z}_t] \leq Z_t \left(1 + \frac{2\beta}{\zeta N} \right)^t. \quad (181)$$

Corollary 7.7. *If Assumptions 4.A, 5.C and 5.D hold, then for all $y \in Y$,*

$$P_\theta^{\alpha,N}(y, dz) \geq \epsilon_{t,N} \pi(dz), \quad (182)$$

where

$$\epsilon_{t,N} \triangleq \frac{(1 - \frac{1}{N})(1 - \kappa_N)^{t-1}}{(1 + \frac{2\beta}{\zeta N})^t}. \quad (183)$$

Furthermore, if $N \geq \frac{2\beta}{Ct(1-\kappa'_N)-\kappa'_N}$ for some constant $C > 0$, where $\kappa'_N \triangleq \kappa_N \vee N^{-1}$, then

$$\epsilon_{t,N} \geq \exp\left(-\frac{1}{\zeta C}\right). \quad (184)$$

In particular, assuming $\kappa'_N \leq B/N$ for some constant $B > 0$, if $N \geq Ct + B$, then

$$\epsilon_{t,N} \geq \exp\left(-\frac{2\beta}{\zeta C} - B\right). \quad (185)$$

Proof. The first part follows from Propositions 7.5 and 7.6. For the second part, we then have

$$\epsilon_{t,N} \geq \left(\frac{1 + \frac{2\beta}{\zeta N}}{1 - \kappa'_N} \right)^{-t} = \left(1 + \frac{1}{1 - \kappa'_N} \left(\frac{2\beta}{\zeta N} + \kappa'_N \right) \right)^{-t} \quad (186)$$

$$\geq \left(1 + \frac{1}{\zeta C t} \right)^{-t} \geq \exp\left(-\frac{1}{\zeta C}\right). \quad (187)$$

The final part follows after noting that if $\kappa'_N \leq B/N$, then

$$\frac{1}{1 - \kappa'_N} \left(\frac{2\beta}{\zeta N} + \kappa'_N \right) \geq \frac{1}{1 - B/N} \left(\frac{2\beta}{\zeta N} + B/N \right) = \frac{1}{N - B} \left(\frac{2\beta}{\zeta} + B \right). \quad (188)$$

□

Remark 7.2. In the case of SIR, Corollary 7.7 is almost as good as [2, Corollary 15]: the former result replaces $\beta - 1$ with β .

The following, which generalizes Corollary 7.3 and Theorem 7.4, is a straightforward consequence of Propositions 7.2 and 5.6 and Corollary 7.7:

Theorem 7.8. *If Assumptions 5.C and 5.D hold, then for $N \geq 2$, the i -caSMC process with kernel $P = P_\theta^{\alpha,N}$*

- (1) is reversible with respect to π and defines a positive operator,
 (2) if the potentials are bounded there exists $\epsilon_{t,N} = 1 + O(1/N)$ such that
 (a) for all $y \in Y$, $P(y, dz) \geq \epsilon_{t,N} \pi_\theta(dz)$,
 (b) for any measure $\nu \ll \pi_\theta$ and $k \geq 1$,

$$d_{\chi^2}(\nu P^k, \pi_\theta) \leq d_{\chi^2}(\nu, \pi_\theta)(1 - \epsilon_{t,N})^k, \quad (189)$$

- (c) for any $y \in Y$ and $k \geq 1$,

$$d_{TV}(\delta_y P^k, \pi_\theta) \leq (1 - \epsilon_{t,N})^k, \quad (190)$$

- (d) for any $\phi \in L^2(Y, \pi_\theta)$

$$\mathbb{V}[\phi, P] \leq (2\epsilon_{t,N}^{-1} - 1)\mathbb{V}_\pi[\phi], \quad (191)$$

- (3) if in addition Assumption 4.A holds and there is a constant $B > 0$ such that $\kappa'_N \leq B/N$, then for every $C > 0$ there exists an $\varepsilon_{B,C,\zeta} > 0$ such that with $N \geq Ct - B$, for any $t > 1$,

$$\epsilon_{t,N} \geq \varepsilon_{B,C,\zeta} > 0. \quad (192)$$

Furthermore, if

$$\pi\text{-ess sup}_\theta \frac{\prod_{s=1}^t \bar{g}_{\theta,s}}{\gamma_\theta(1)} < \infty \quad (193)$$

or there exists $1 \leq \beta < \infty$ such that for any $t, s \in \mathbb{N}$,

$$\pi\text{-ess sup}_{\theta,x} \frac{G_{0,t} G_{t,t+s}(x)}{G_{0,t+s}} \leq \beta. \quad (194)$$

then, when Assumptions 5.C and 5.D hold and the Gibbs sampler is geometrically ergodic, the α PG chain is geometrically ergodic.

Notably, our results show that ∞ -ESS is a notion of effective sample size in the setting of the i- α SMC and α PG algorithms.

Remark 7.3. At a high level, we have seen that the expected value of \tilde{Z}_t arises in the study of the mixing properties of iterated conditional SMC (i-cSMC) Markov chains, which are related to the convergence of particle Gibbs (PG) samplers. In order to show geometric ergodicity for particle Gibbs samplers, bounds on the expected value of \tilde{Z}_t can be used, with growth of the expectation as t increases determining how well the particle Gibbs algorithm scales. Bounds on the expected value of \tilde{Z}_t also allow one to obtain bounds on $\text{KL}(\pi_{1,t} || \bar{\pi}_{1,t}^{R,N})$.² We have also shown that an analogous connection exists for the expected value of \tilde{Z}_t between adaptive particle Gibbs and adaptive SMC for sampling. Hence, as a slogan, good performance of (adaptive) particle Gibbs is equivalent to good performance of (adaptive) SMC for sampling.

²In the i-cSMC setting, the expectation is with respect to the “doubly conditional SMC kernel,” whereas we require expectations with respect to the “conditional SMC kernel.” However, this is only a small technical difference and, as we have seen, the same techniques apply in both cases.

APPENDIX A. ADDITIONAL PROOFS

A.1. Proof of Proposition 5.3. For (1), the fact that $\mathcal{E}_t^1 = \mathcal{E}_t^{ent}$ is a straightforward algebraic manipulation. To prove the limit equality, write $\mathbf{v} = \langle v_1, \dots, v_N \rangle$ and observe that, using the Taylor series for x^p and $\log(1+x)$, we have

$$\lim_{p \rightarrow 1} \left(\frac{\|\mathbf{v}\|_1^p}{\sum_{n=1}^N v_n^p} \right)^{1/(1-p)} \quad (195)$$

$$= \lim_{p \rightarrow 1} \left(\frac{\sum_{k=0}^{\infty} \|\mathbf{v}\|_1 (p-1)^k \log^k(\|\mathbf{v}\|_1)/k!}{\sum_{n=1}^N \sum_{k=0}^{\infty} v_n (p-1)^k \log^k(v_n)/k!} \right)^{1/(1-p)} \quad (196)$$

$$= \lim_{x \rightarrow \infty} \left(\frac{\sum_{k=0}^{\infty} \|\mathbf{v}\|_1 x^{-k} \log^k(\|\mathbf{v}\|_1)/k!}{\sum_{n=1}^N \sum_{k=0}^{\infty} v_n x^{-k} \log^k(v_n)/k!} \right)^x \quad (197)$$

$$= \lim_{x \rightarrow \infty} \left(\frac{\exp(\log(1 + \sum_{k=1}^{\infty} x^{-k} \log^k(\|\mathbf{v}\|_1)/k!))}{\exp(\log(1 + \sum_{k=1}^{\infty} \sum_{n=1}^N v_n \|\mathbf{v}\|_1^{-1} x^{-k} \log^k(v_n)/k!))} \right)^x \quad (198)$$

$$= \lim_{x \rightarrow \infty} \frac{\exp(x \sum_{m=1}^{\infty} (-1)^{m+1} [\sum_{k=1}^{\infty} x^{-k} \log^k(\|\mathbf{v}\|_1)/k!]^m)}{\exp(x \sum_{m=1}^{\infty} (-1)^{m+1} [\sum_{k=1}^{\infty} \sum_{n=1}^N v_n \|\mathbf{v}\|_1^{-1} x^{-k} \log^k(v_n)/k!]^m)} \quad (199)$$

$$= \lim_{x \rightarrow \infty} \frac{\exp(\log(\|\mathbf{v}\|_1) + \Theta(x^{-1}))}{\exp(\log(\prod_{n=1}^N v_n^{v_n/\|\mathbf{v}\|_1}) + \Theta(x^{-1}))} \quad (200)$$

$$= \frac{\|\mathbf{v}\|_1}{\prod_{n=1}^N v_n^{v_n/\|\mathbf{v}\|_1}}. \quad (201)$$

To prove the remaining parts, we make repeated use of the following:

Fact. For $1 \leq r < s \leq \infty$, and any vector $\mathbf{v} \in \mathbb{R}_+^N$, $\|\mathbf{v}\|_s \leq \|\mathbf{v}\|_r \leq N^{1/r-1/s} \|\mathbf{v}\|_s$, with the lower (upper) bound achieved if and only if \mathbf{v} has one non-zero entry (\mathbf{v} has all equal entries).

For (2), apply the Fact with $r = 1$, $s = p > 1$, and note that in this case $1/r - 1/s = 1 - 1/p = 1/p_*$. We then have $1 \leq \|\mathbf{v}\|_1/\|\mathbf{v}\|_p \leq N^{1/p_*}$, proving the result for $p > 1$. For $p = 1$, the result follows from part (1) and elementary properties of the entropy.

For (3), in the case that $p > 1$, note that

$$\begin{aligned} \|\mathbf{v}\|_1^{q_*-p_*} &\geq N^{(q_*-p_*)/q_*} \|\mathbf{v}\|_q^{q_*-p_*} = N^{1-p_*/q_*} \|\mathbf{v}\|_q^{q_*-p_*} \\ &= N^{-p_*(1/p-1/q)} \|\mathbf{v}\|_q^{q_*-p_*}, \end{aligned} \quad (202)$$

where the final equality follows since

$$1 - p_*/q_* = 1 - p_*(1 - 1/q) = 1 - p_* + p_*/q = -p_*/p + p_*/q. \quad (203)$$

We conclude that

$$\left(\frac{\|\mathbf{v}\|_1}{\|\mathbf{v}\|_p} \right)^{p_*} \geq \frac{\|\mathbf{v}\|_1^{p_*}}{N^{p_*(1/p-1/q)} \|\mathbf{v}\|_q^{p_*}} \quad (204)$$

$$\geq \frac{\|\mathbf{v}\|_1^{p_*}}{N^{p_*(1/p-1/q)} \|\mathbf{v}\|_q^{p_*}} \frac{\|\mathbf{v}\|_1^{q_*-p_*}}{N^{-p_*(1/p-1/q)} \|\mathbf{v}\|_q^{q_*-p_*}} \quad (205)$$

$$= \left(\frac{\|\mathbf{v}\|_1}{\|\mathbf{v}\|_q} \right)^{q_*} \quad (206)$$

$$\geq \left(\frac{\|\mathbf{v}\|_1}{\|\mathbf{v}\|_p} \right)^{q_*} \quad (207)$$

$$= \left(\frac{\|\mathbf{v}\|_p}{\|\mathbf{v}\|_1} \right)^{p_* - q_*} \left(\frac{\|\mathbf{v}\|_1}{\|\mathbf{v}\|_p} \right)^{p_*} \quad (208)$$

$$\geq N^{-(p_* - q_*)/p_*} \left(\frac{\|\mathbf{v}\|_1}{\|\mathbf{v}\|_p} \right)^{p_*}, \quad (209)$$

where the first, third, and fourth inequalities follow from the Fact and the second follows from Eq. (202).

The case of $p = 1$ follows from the $p > 1$ case and part (1).

A.2. Proof of Proposition 5.4. We prove the result for $i = 1$. The general case follows from straightforward modifications. We will use the abbreviated notation $Q_{s,t}^k(\cdot) = Q_{s,t}(\cdot)(X_s^k)$ or $Q_{s,t}(\cdot)(x_s^k)$, $G_{s,t}^k = G_{s,t}(X_s^k)$ or $G_{s,t}(x_s^k)$, $G_{s,t}^y = G_{s,t}(y_s)$, $g_s^k = g_s(X_s^k)$ or $g_s(x_s^k)$, and $g_s^y = g_s(y_s)$. The variables are X_s^k inside expectations and x_s^k outside expectations. Throughout the proof, when limits of a sum are not specified, the sum is from 1 to N .

The proof relies on the following lemma.

Lemma A.1. *If $y_{1,t} \in E^t$, then*

(1) *for $s = 2, \dots, t$ and any functions $\phi_s^n : E \rightarrow \mathbb{R}$, $n \in [N]$,*

$$\begin{aligned} & \mathbb{E}_{y_{1,t}}^{\alpha, N} \left[\sum_n \phi_s^n(X_s^n) \mid \mathcal{F}_{s-1} \right] \\ &= \sum_{f_s} \alpha_{s-1}^{f_s f_{s-1}} \phi_s^{f_s}(y_s) + \sum_{f_s} \sum_{n \neq f_s} \sum_k \alpha_{s-1}^{f_s f_{s-1}} \frac{\alpha_{s-1}^{nk} w_{s-1}^k}{w_s^n} Q_{s-1,s}^k(\phi_s^n); \end{aligned} \quad (210)$$

(2) *for $\tau \in [t - s]$,*

$$\begin{aligned} & \mathbb{E}_{y_{1,t}}^{\alpha, N} \left[\sum_n W_s^n G_{s,s+\tau}^n \mid \mathcal{F}_{s-1} \right] \\ & \leq \frac{1}{\zeta N} \sum_n w_{s-1}^n g_{s-1}^n G_{s,s+\tau}^y + \sum_n w_{s-1}^n G_{s-1,s+\tau}^n; \end{aligned} \quad (211)$$

and

(3) *for $s = 1, \dots, t - 1$,*

$$N \mathbb{E}_{y_{1,t}}^{\alpha, N} [\ddot{Z}_t \mid \mathcal{F}_{t-s}] \leq A_{t-s} + B_{t-s}, \quad (212)$$

where,

$$A_{t-s} \triangleq (\zeta N)^{-s+1} \sum_n w_{t-s}^n g_{t-s}^n \left(\sum_{\ell=1}^s \sum_{\tau \in \mathcal{T}_{\ell,s}} (\zeta N)^{s-1-\ell} G_{t-s+1,\tau_1}^y C_\ell^y(\tau) \right) \quad (213)$$

$$B_{t-s} \triangleq (\zeta N)^{-s+1} \sum_n w_{t-s}^n \left(\sum_{\ell=1}^s \sum_{\tau \in \mathcal{T}_{\ell,s}} (\zeta N)^{s-\ell} G_{t-s,\tau_1}^n C_\ell^y(\tau) \right). \quad (214)$$

Proof. For (1),

$$\begin{aligned}
& \mathbb{E}_{y_{1,t}}^{\alpha,N} \left[\sum_n \phi_s^n(X_s^n) \mid \mathcal{F}_{s-1} \right] \\
&= \sum_{f_s} \sum_{a_{s-1}^{f_s}} \alpha_{s-1}^{f_s f_{s-1}} \prod_{k \neq f_s} r_k(a_{s-1}^k \mid \mathbf{w}_{s-1}, \mathbf{x}_{1,s-1}) \mathbb{E}_{y_{1,t}}^{\alpha,N} \left[\sum_n \phi_s^n(X_s^n) \mid \mathcal{F}_{s-1}, \mathbf{A}_{s-1} = \mathbf{a}_{s-1}, F_s = f_s \right] \\
&= \sum_{f_s} \sum_{a_{s-1}^{f_s}} \alpha_{s-1}^{f_s f_{s-1}} \prod_{k \neq f_s} \frac{\alpha_{s-1}^{k a_{s-1}^k} w_{s-1}^{a_{s-1}^k} g_{s-1}(x_{s-1}^{a_{s-1}^k})}{w_s^k} \left(\phi_s^{f_s}(y_s) + \sum_{n \neq f_s} \mathbb{E} \left[\phi_s^n(X_s) \mid X_{s-1} = x_{s-1}^{a_{s-1}^n} \right] \right) \\
&= \sum_{f_s} \alpha_{s-1}^{f_s f_{s-1}} \phi_s^{f_s}(y_s) + \sum_{f_s} \sum_{n \neq f_s} \sum_k \alpha_{s-1}^{f_s f_{s-1}} \frac{\alpha_{s-1}^{nk} w_{s-1}^k g_{s-1}(x_{s-1}^k)}{w_s^n} \mathbb{E} [\phi_s^n(X_s) \mid X_{s-1} = x_{s-1}^k] \\
&= \sum_{f_s} \alpha_{s-1}^{f_s f_{s-1}} \phi_s^{f_s}(y_s) + \sum_{f_s} \sum_{n \neq f_s} \sum_k \alpha_{s-1}^{f_s f_{s-1}} \frac{\alpha_{s-1}^{nk} w_{s-1}^k}{w_s^n} Q_{s-1,s}^k(\phi_s^n)
\end{aligned}$$

For (2), choosing $\phi_s^n(x) = w_s^n G_{s,s+\tau}(x)$, we have

$$\mathbb{E}_{y_{1,t}}^{\alpha,N} \left[\sum_n W_s^n G_{s,s+\tau}^n \mid \mathcal{F}_{s-1} \right] \quad (215)$$

$$= \sum_{f_s} \alpha_{s-1}^{f_s f_{s-1}} w_s^{f_s} G_{s,s+\tau}^{f_s} + \sum_{f_s} \sum_{n \neq f_s} \sum_k \alpha_{s-1}^{f_s f_{s-1}} \frac{\alpha_{s-1}^{nk} w_{s-1}^k}{w_s^n} Q_{s-1,s}^k(w_s^n G_{s,s+\tau}) \quad (216)$$

$$= G_{s,s+\tau}^{f_s} \sum_{f_s} \alpha_{s-1}^{f_s f_{s-1}} w_s^{f_s} + \sum_{f_s} \sum_{n \neq f_s} \sum_k \alpha_{s-1}^{f_s f_{s-1}} \alpha_{s-1}^{nk} w_{s-1}^k G_{s-1,s+\tau}^k \quad (217)$$

$$\leq G_{s,s+\tau}^{f_s} \sum_{f_s} \alpha_{s-1}^{f_s f_{s-1}} \frac{\|\mathbf{w}_s\|_1}{\zeta N} + \sum_{f_s} \sum_n \sum_k \alpha_{s-1}^{f_s f_{s-1}} \alpha_{s-1}^{nk} w_{s-1}^k G_{s-1,s+\tau}^k \quad (218)$$

$$= \frac{G_{s,s+\tau}^{f_s}}{\zeta N} \sum_n w_s^n + \sum_n \sum_k \alpha_{s-1}^{nk} w_{s-1}^k G_{s-1,s+\tau}^k \quad (219)$$

$$= \frac{G_{s,s+\tau}^{f_s}}{\zeta N} \sum_n \sum_k \alpha_{s-1}^{nk} w_{s-1}^k g_{s-1}^k + \sum_k w_{s-1}^k G_{s-1,s+\tau}^k \quad (220)$$

$$= \frac{1}{\zeta N} \sum_k w_{s-1}^k g_{s-1}^k G_{s,s+\tau}^{f_s} + \sum_k w_{s-1}^k G_{s-1,s+\tau}^k, \quad (221)$$

where the inequality follows from Assumption 5.D, and we have repeatedly used Assumption 5.C.

To show (3), we start by using (2) with $s = t$ and $\tau = 1$:

$$\mathbb{E}_{y_{1,t}}^{\alpha,N} \left[\sum_n W_t^n g_t^n \mid \mathcal{F}_{t-1} \right] = \frac{1}{\zeta N} \sum_k w_{t-1}^k g_{t-1}^k g_t^{f_t} + \frac{\zeta N}{\zeta N} \sum_m w_{t-1}^m G_{t-1,t+1}^m \quad (222)$$

$$= A_{t-1} + B_{t-1}, \quad (223)$$

Hence, (3) holds for $s = 1$. We now assume that the bound holds for some $s \in \{1, \dots, t-2\}$ and establish that it also holds for $s+1$. Using the inductive

hypothesis,

$$N \mathbb{E}_{y_{1,t}}^{\alpha,N} [\ddot{Z}_t | \mathcal{F}_{t-s-1}] = \mathbb{E}_{y_{1,t}}^{\alpha,N} [N \mathbb{E}_{y_{1,t}}^{\alpha,N} [\ddot{Z}_t | \mathcal{F}_{t-s}] | \mathcal{F}_{t-s-1}] \quad (224)$$

$$\leq \mathbb{E}_{y_{1,t}}^{\alpha,N} [A_{t-s} + B_{t-s} | \mathcal{F}_{t-s-1}]. \quad (225)$$

Using (2), we have

$$A \triangleq \mathbb{E}_{y_{1,t}}^{\alpha,N} [A_{t-s} | \mathcal{F}_{t-s-1}] \quad (226)$$

$$= (\zeta N)^{-s+1} \left(\sum_{\ell=1}^s \sum_{\tau \in \mathcal{T}_{\ell,s}} (\zeta N)^{s-1-\ell} G_{t-s+1,\tau_1}^y C_{\ell}^y(\tau) \right) \mathbb{E}_{y_{1,t}}^{\alpha,N} \left[\sum_n W_{t-s}^n g_{t-s}^n \right] \quad (227)$$

$$\leq (\zeta N)^{-s} \left(\sum_{\ell=1}^s \sum_{\tau \in \mathcal{T}_{\ell,s}} (\zeta N)^{s-1-\ell} G_{t-s+1,\tau_1}^y C_{\ell}^y(\tau) \right) \times \left(\sum_n w_{t-s-1}^n g_{t-s-1}^n g_{t-s}^y + \zeta N \sum_n w_{t-s-1}^n G_{t-s-1,t-s+1}^n \right) \quad (228)$$

and

$$B \triangleq \mathbb{E}_{y_{1,t}}^{\alpha,N} [B_{t-s} | \mathcal{F}_{t-s-1}] \quad (229)$$

$$= (\zeta N)^{-s+1} \left(\sum_{\ell=1}^s \sum_{\tau \in \mathcal{T}_{\ell,s}} (\zeta N)^{s-\ell} \mathbb{E}_{y_{1,t}}^{\alpha,N} \left[\sum_n W_{t-s}^n G_{t-s,\tau_1}^n | \mathcal{F}_{t-s-1} \right] C_{\ell}^y(\tau) \right) \quad (230)$$

$$\leq (\zeta N)^{-s} \left(\sum_{\ell=1}^s \sum_{\tau \in \mathcal{T}_{\ell,s}} (\zeta N)^{s-\ell} \left(\sum_n w_{t-s-1}^n g_{t-s-1}^n G_{t-s,\tau_1}^y + \zeta N \sum_n w_{t-s-1}^n G_{t-s-1,\tau_1}^n \right) C_{\ell}^y(\tau) \right). \quad (231)$$

Hence,

$$\begin{aligned} A + B &\leq (\zeta N)^{-s} \sum_n w_{t-s-1}^n g_{t-s-1}^n \left(\sum_{\ell=1}^s \sum_{\tau \in \mathcal{T}_{\ell,s}} (\zeta N)^{s-1-\ell} G_{t-s,t-s+1}^y G_{t-s+1,\tau_1}^y C_{\ell}^y(\tau) \right) \\ &\quad + (\zeta N)^{-s} \sum_n w_{t-s-1}^n g_{t-s-1}^n \left(\sum_{\ell=1}^s \sum_{\tau \in \mathcal{T}_{\ell,s}} (\zeta N)^{s-\ell} G_{t-s,\tau_1}^y C_{\ell}^y(\tau) \right) \\ &\quad + (\zeta N)^{-s} \sum_n w_{t-s-1}^n \left(\sum_{\ell=1}^s \sum_{\tau \in \mathcal{T}_{\ell,s}} (\zeta N)^{s-\ell} G_{t-s-1,t-s+1}^n G_{t-s+1,\tau_1}^y C_{\ell}^y(\tau) \right) \\ &\quad + (\zeta N)^{-s} \sum_n w_{t-s-1}^n \left(\sum_{\ell=1}^s \sum_{\tau \in \mathcal{T}_{\ell,s}} (\zeta N)^{s-\ell+1} G_{t-s-1,\tau_1}^n C_{\ell}^y(\tau) \right). \end{aligned} \quad (232)$$

Summing the parenthesized double sums of the first two terms yields

$$\begin{aligned} & \sum_{\ell=1}^s \sum_{\tau \in \mathcal{T}_{\ell,s}} (\zeta N)^{s-1-\ell} G_{t-s,t-s+1}^y G_{t-s+1,\tau_1}^y C_{\ell}^y(\tau) + \sum_{\ell=1}^s \sum_{\tau \in \mathcal{T}_{\ell,s}} (\zeta N)^{s-\ell} G_{t-s,\tau_1}^y C_{\ell}^y(\tau) \\ &= \sum_{\ell=1}^{s+1} \sum_{\substack{\tau \in \mathcal{T}_{\ell,s+1} \\ \tau_1 = t-s+1}} (\zeta N)^{s-\ell} G_{t-s,\tau_1}^y C_{\ell}^y(\tau) + \sum_{\ell=1}^{s+1} \sum_{\substack{\tau \in \mathcal{T}_{\ell,s+1} \\ \tau_1 > t-s+1}} (\zeta N)^{s-\ell} G_{t-s,\tau_1}^y C_{\ell}^y(\tau) \quad (233) \end{aligned}$$

$$= \sum_{\ell=1}^{s+1} \sum_{\tau \in \mathcal{T}_{\ell,s+1}} (\zeta N)^{s-\ell} G_{t-s,\tau_1}^y C_{\ell}^y(\tau), \quad (234)$$

so the first two terms are equal to $A_{t-(s+1)}$. Summing the parenthesized double sums of the last two terms yields

$$\begin{aligned} & \sum_{\ell=1}^s \sum_{\tau \in \mathcal{T}_{\ell,s}} (\zeta N)^{s-\ell} G_{t-s-1,t-s+1}^n G_{t-s+1,\tau_1}^y C_{\ell}^y(\tau) + \sum_{\ell=1}^s \sum_{\tau \in \mathcal{T}_{\ell,s}} (\zeta N)^{s-\ell+1} G_{t-s-1,\tau_1}^n C_{\ell}^y(\tau) \\ &= \sum_{\ell=1}^{s+1} \sum_{\substack{\tau \in \mathcal{T}_{\ell,s+1} \\ \tau_1 = t-s+1}} (\zeta N)^{s-\ell+1} G_{t-s-1,\tau_1}^n C_{\ell}^y(\tau) + \sum_{\ell=1}^{s+1} \sum_{\substack{\tau \in \mathcal{T}_{\ell,s+1} \\ \tau_1 > t-s+1}} (\zeta N)^{s-\ell+1} G_{t-s-1,\tau_1}^n C_{\ell}^y(\tau) \quad (235) \end{aligned}$$

$$= \sum_{\ell=1}^{s+1} \sum_{\tau \in \mathcal{T}_{\ell,s+1}} (\zeta N)^{s-\ell+1} G_{t-s-1,\tau_1}^n C_{\ell}^y(\tau), \quad (236)$$

so the last two terms are equal to $B_{t-(s+1)}$. \square

Using part (3) of Lemma A.1 with $s = t - 1$, we have

$$N \mathbb{E}_{y_{1,t}}^{\alpha,N} [\ddot{Z}_t] \leq \mathbb{E}_{y_{1,t}}^{\alpha,N} [A_1 + B_1]. \quad (237)$$

Therefore,

$$\mathbb{E}_{y_{1,t}}^{\alpha,N} [A_1] = (\zeta N)^{-t+2} \mathbb{E}_{y_{1,t}}^{\alpha,N} \left[\sum_n g_1^n \right] \left(\sum_{\ell=1}^{t-1} \sum_{\tau \in \mathcal{T}_{\ell,t-1}} (\zeta N)^{t-2-\ell} G_{2,\tau_1}^y C_{\ell}^y(\tau) \right) \quad (238)$$

$$= (\zeta N)^{-t+2} (G_{1,2}^y + (N-1)G_{0,2}) \left(\sum_{\ell=1}^{t-1} \sum_{\tau \in \mathcal{T}'_{\ell,t-1}} (\zeta N)^{t-2-\ell} G_{2,\tau_1}^y C_{\ell}^y(\tau) \right) \quad (239)$$

and

$$\mathbb{E}_{y_{1,t}}^{\alpha,N} [B_1] = (\zeta N)^{-t+2} \left(\sum_{\ell=1}^{t-1} \sum_{\tau \in \mathcal{T}_{\ell,t-1}} (\zeta N)^{t-1-\ell} \mathbb{E}_{y_{1,t}}^{\alpha,N} \left[\sum_n G_{1,\tau_1}^n \right] C_{\ell}^y(\tau) \right) \quad (240)$$

$$= (\zeta N)^{-t+2} \left(\sum_{\ell=1}^{t-1} \sum_{\tau \in \mathcal{T}_{\ell,t-1}} (\zeta N)^{t-1-\ell} (G_{1,\tau_1}^y + (N-1)G_{0,\tau_1}) C_{\ell}^y(\tau) \right). \quad (241)$$

Hence, using arguments analogous to those from the proof of Lemma A.1 and the fact that $G_{0,1} = 1$ yields

$$\begin{aligned} & \mathbb{E}_{y_{1,t}}^{\alpha,N} [A_1 + B_1] \\ &= (\zeta N)^{-t+2} \sum_{\ell=1}^t \sum_{\boldsymbol{\tau} \in \mathcal{T}_{\ell,t}} (\zeta N)^{t-1-\ell} G_{1,\tau_1}^y C_{\ell}^y(\boldsymbol{\tau}) \end{aligned} \quad (242)$$

$$+ \frac{N-1}{\zeta N} (\zeta N)^{-t+2} \sum_{\ell=1}^t \sum_{\boldsymbol{\tau} \in \mathcal{T}_{\ell,t}} (\zeta N)^{t-\ell} G_{0,\tau_1} C_{\ell}^y(\boldsymbol{\tau}) \quad (243)$$

$$\begin{aligned} &= (\zeta N)^{-t+2} \sum_{\ell=1}^{t+1} \sum_{\substack{\boldsymbol{\tau} \in \mathcal{T}_{\ell,t+1} \\ \tau_1=1}} (\zeta N)^{t-\ell} G_{0,\tau_1} C_{\ell}^y(\boldsymbol{\tau}) \\ &+ \frac{N-1}{\zeta N} (\zeta N)^{-t+2} \sum_{\ell=1}^{t+1} \sum_{\substack{\boldsymbol{\tau} \in \mathcal{T}_{\ell,t+1} \\ \tau_1>1}} (\zeta N)^{t-\ell} G_{0,\tau_1} C_{\ell}^y(\boldsymbol{\tau}) \end{aligned} \quad (244)$$

$$= (\zeta N)^{-t+2} \sum_{\ell=1}^{t+1} \sum_{\boldsymbol{\tau} \in \mathcal{T}_{\ell,t+1}} (\zeta N)^{t-\ell} \left(\frac{N-1}{\zeta N} \right)^{\mathbb{1}(\tau_1>1)} G_{0,\tau_1} C_{\ell}^y(\boldsymbol{\tau}). \quad (245)$$

A.3. Proof of Proposition 7.5. First, observe that we can write the i- α SMC kernel as

$$P_{\theta}^{\alpha,N}(y, dz) = \mathbb{E}_{y,\theta}^{\alpha,N} \left[\sum_{\mathbf{k} \in [N]^t} \mathcal{I}_{\mathbf{k}}^{\alpha}(\mathbf{X}_{1,t}, \mathbf{A}_{1,t}, dz) \right], \quad (246)$$

where

$$\mathcal{I}_{\mathbf{k}}^{\alpha}(\mathbf{x}_{1,t}, \mathbf{a}_{1,t}, dz) \triangleq \delta_{x_{1,t}^{k_t}}(dz) \mathbb{1}(k_t = a_t) \prod_{s=1}^{t-1} \mathbb{1}(k_s = a_s^{k_{s+1}}). \quad (247)$$

Next note that

$$\begin{aligned} & \sum_{k_1=1}^N \mathbb{1}(x_{1,t}^{\mathbf{k}} \in S) \mathbb{P}_{y,\theta}^{\alpha,N} [\mathbf{X}_1 \in d\mathbf{x}_1, F_1^y = F_1^y] \\ &= \sum_{k_1=1}^N \mathbb{1}(x_{1,t}^{\mathbf{k}} \in S) \frac{1}{N} \delta_{y_1}(dx_1^{f_1^y}) \prod_{n \neq f_1}^N M_1(dx_1^n) \\ &\geq \frac{N}{C_1} \sum_{k_1=1}^N \int_E \mathbb{1}(x_{1,t}^{\mathbf{k}} \in S) \frac{C_1}{N^2} \mathbb{1}(f_1^y \neq k_1) \delta_{y_1}(dx_1^{f_1^y}) \delta_{z_1}(dx_1^{k_1}) \prod_{n \neq f_1^y, k_1}^N M_1(dx_1^n) M_1(dz_1) \\ &= \frac{N}{C_1} \sum_{k_1=1}^N \int_E \mathbb{1}(x_{1,t}^{\mathbf{k}} \in S) \mathbb{P}_{y,z,\theta}^{\alpha,N} [\mathbf{X}_1 \in d\mathbf{x}_1, F_1^y = f_1^y, F_1^z = k_1] M_1(dz_1). \end{aligned}$$

For the remainder of the proof, to keep notation compact when writing laws, instead of writing, e.g., $\mathbf{X}_s \in \mathbf{x}_s$ or $F_s^y = f_s^f$, whenever a random variable is instantiated to be (the differential) of the lowercase version of itself, we will write only the random

variance: for example, \mathbf{X}_s or F_s^y . Now, for $s = 2, \dots, t$,

$$\begin{aligned} & \sum_{k_s=1}^N \mathbb{1}(x_{1,t}^{\mathbf{k}} \in S) \mathbb{1}(k_{s-1} = a_{s-1}^{k_s}) \mathbb{P}_{y,\theta}^{\alpha,N}[\mathbf{X}_s, \mathbf{A}_{s-1}, F_s^y \mid \mathbf{X}_{1,s-1}, \mathbf{A}_{1,s-2}, F_{s-1}^y] \\ &= \sum_{k_s=1}^N \mathbb{1}(x_{1,t}^{\mathbf{k}} \in S) \mathbb{1}(k_{s-1} = a_{s-1}^{k_s}) \alpha_{s-1}^{f_s^y f_{s-1}^y} \mathbb{1}(a_{s-1}^{f_s^y} = f_{s-1}^y) \delta_{y_s}(\mathrm{d}x_s^{f_s^y}) \end{aligned} \quad (248)$$

$$\begin{aligned} & \times \prod_{n \neq f_s^y} r_n(a_{s-1}^n | \mathbf{w}_{s-1}, \mathbf{x}_{1,s-1}) M_s(x_{s-1}^{a_{s-1}^n}, x_s^n) \\ & \geq \sum_{k_s=1}^N \frac{1}{C_s \alpha_{s-1}^{k_s k_{s-1}}} \int_E \mathbb{1}(x_{1,t}^{\mathbf{k}} \in S) C_s \mathbb{1}(f_s^y \neq k_s) \mathbb{1}(a_{s-1}^{f_s^y} = f_{s-1}^y) \mathbb{1}(a_{s-1}^{k_s} = k_{s-1}) \\ & \quad \times \alpha_{s-1}^{f_s^y f_{s-1}^y} \alpha_{s-1}^{k_s k_{s-1}} \delta_{y_s}(\mathrm{d}x_s^{f_s^y}) \delta_{z_s}(\mathrm{d}x_s^{k_s}) \\ & \quad \times \prod_{n \neq f_s^y, k_s} r_n(a_{s-1}^n | \mathbf{w}_{s-1}, \mathbf{x}_{1,s-1}) M_s(x_{s-1}^{a_{s-1}^n}, x_s^n) \\ & \quad \times M_s(x_{s-1}^{k_{s-1}}, \mathrm{d}z_s) r_{k_s}(k_{s-1} | \mathbf{w}_{s-1}, \mathbf{x}_{1,s-1}) \end{aligned} \quad (249)$$

$$\begin{aligned} &= \sum_{k_s=1}^N \frac{1}{C_s \alpha_{s-1}^{k_s k_{s-1}}} \int_E \mathbb{1}(x_{1,t}^{\mathbf{k}} \in S) M_s(x_{s-1}^{k_{s-1}}, \mathrm{d}z_s) r_{k_s}(k_{s-1} | \mathbf{w}_{s-1}, \mathbf{x}_{1,s-1}) \\ & \quad \times \mathbb{P}_{y,z,\theta}^{\alpha,N}[\mathbf{X}_s, \mathbf{A}_{s-1}, F_s^y, F_s^z = k_s \mid \mathbf{X}_{1,s-1}, \mathbf{A}_{1,s-2}, F_{s-1}^y, F_{s-1}^z = k_{s-1}]. \end{aligned} \quad (250)$$

Using Eqs. (248) and (250), we have (note that the terms such as those involving a_0 should be ignored)

$$\begin{aligned} & \sum_{\mathbf{k} \in [N]^t} \mathbb{1}(x_{1,t}^{\mathbf{k}} \in S) \mathbb{1}(k_t = a_t) \prod_{s=1}^t \mathbb{1}(k_{s-1} = a_{s-1}^{k_s}) \mathbb{P}_{y,\theta}^{\alpha,N}[\mathbf{X}_{1,t}, \mathbf{A}_{1,t}, \mathbf{F}_{1,t}^y] \\ &= \sum_{\mathbf{k} \in [N]^t} \mathbb{1}(x_{1,t}^{\mathbf{k}} \in S) \mathbb{1}(k_t = a_t) \mathbb{P}_{y,\theta}^{\alpha,N}[A_t \mid \mathbf{X}_{1,t}, \mathbf{A}_{1,t-1}] \end{aligned} \quad (251)$$

$$\begin{aligned} & \times \prod_{s=1}^t \mathbb{1}(k_{s-1} = a_{s-1}^{k_s}) \mathbb{P}_{y,\theta}^{\alpha,N}[\mathbf{X}_s, \mathbf{A}_{s-1}, F_s^y \mid \mathbf{X}_{1,s-1}, \mathbf{A}_{1,s-2}, F_{s-1}^y] \\ & \geq \sum_{\mathbf{k} \in [N]^t} \int_Y \frac{N \mathbb{1}(x_{1,t}^{\mathbf{k}} \in S) \mathbb{1}(k_t = a_t)}{(\prod_{s=1}^t C_s) (\prod_{s=1}^{t-1} \alpha_{s-1}^{k_s k_{s-1}})} \frac{w_t^{a_t} g_t(x_t^{a_t})}{\sum_{n=1}^N w_t^n g_t(x_t^n)} \\ & \quad \times \prod_{s=1}^t M_s(x_{s-1}^{k_{s-1}}, \mathrm{d}z_s) r_{k_s}(k_{s-1} | \mathbf{w}_{s-1}, \mathbf{x}_{1,s-1}) \\ & \quad \times \prod_{s=1}^t \mathbb{P}_{y,z,\theta}^{\alpha,N}[\mathbf{X}_s, \mathbf{A}_{s-1}, F_s^y, F_s^z = k_s \mid \mathbf{X}_{1,s-1}, \mathbf{A}_{1,s-2}, F_{s-1}^y, F_{s-1}^z = k_{s-1}] \end{aligned} \quad (252)$$

$$= \sum_{\mathbf{k} \in [N]^t} \int_Y \frac{N \mathbb{1}(z_{1,t} \in S)}{(\prod_{s=1}^t C_s) \sum_{n=1}^N w_t^n g_t(x_t^n)} \mathbb{P}_{y,z,\theta}^{\alpha,N}[\mathbf{X}_{1,t}, \mathbf{A}_{1,t-1}, \mathbf{F}_{1,t}^y, \mathbf{F}_{1,t}^z = \mathbf{k}] Q_{0,t+1}(dz_{1,t}) \quad (253)$$

$$= \int_S \sum_{\mathbf{k} \in [N]^t} \frac{Z_t}{\ddot{Z}_t \prod_{s=1}^t C_s} \mathbb{P}_{y,z,\theta}^{\alpha,N}[\mathbf{X}_{1,t}, \mathbf{A}_{1,t-1}, \mathbf{F}_{1,t}^y, \mathbf{F}_{1,t}^z = \mathbf{k}] \pi(dz_{1,t}), \quad (254)$$

from which the result follows.

ACKNOWLEDGMENTS

Thanks to Arnaud Doucet for critical feedback and numerous helpful suggestions, to Cameron Freer for feedback on early versions of this work, to Josh Tenenbaum for discussions that helped to inspire this work, and to Vikash Mansinghka for suggesting we investigate the expected value of SMC estimators. JHH was supported by the U.S. Government under FA9550-11-C-0028 and awarded by the DoD, Air Force Office of Scientific Research, National Defense Science and Engineering Graduate (NDSEG) Fellowship, 32 CFR 168a. This research was carried out in part while DMR held a Research Fellowship at Emmanuel College, Cambridge, with funding also from a Newton International Fellowship through the Royal Society.

REFERENCES

- [1] C. Andrieu, A. Doucet, and R. Holenstein. “Particle Markov chain Monte Carlo methods”. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 72.3 (2010), pp. 269–342.
- [2] C. Andrieu, A. Lee, and M. Vihola. “Uniform Ergodicity of the Iterated Conditional SMC and Geometric Ergodicity of Particle Gibbs samplers”. In: *arXiv.org* (Dec. 2013). arXiv:1312.6432v1 [math.PR].
- [3] C. Andrieu and G. O. Roberts. “The pseudo-marginal approach for efficient Monte Carlo computations”. In: *The Annals of Statistics* 37.2 (Apr. 2009), pp. 697–725.
- [4] C. Andrieu and M. Vihola. “Convergence properties of pseudo-marginal Markov chain Monte Carlo algorithms”. In: *The Annals of Applied Probability* (2014).
- [5] N. Chopin and S. S. Singh. “On the particle Gibbs sampler”. In: *arXiv.org* (Apr. 2013). arXiv:1304.1887v1 [stat.CO].
- [6] J. Cornebise, E. Moulines, and J. Olsson. “Adaptive methods for sequential importance sampling with application to state space models”. In: *Statistics and Computing* 18.4 (2008), pp. 461–480.
- [7] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. John Wiley & Sons, 1991.
- [8] D. Crisan and O. Obanubi. “Particle filters with random resampling times”. In: *Stochastic processes and their applications* 122.4 (2012), pp. 1332–1368.
- [9] P. Del Moral. *Feynman-Kac Formulae: Genealogical and Interacting Particles Systems with Applications*. New York: Springer, 2004.
- [10] P. Del Moral, A. Doucet, and A. Jasra. “On adaptive resampling strategies for sequential Monte Carlo methods”. In: *Bernoulli* 18.1 (Feb. 2012), pp. 252–278.
- [11] P. Del Moral, A. Doucet, and A. Jasra. “Sequential Monte Carlo samplers”. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 68.3 (2006), pp. 411–436.

- [12] A. Doucet, N. de Freitas, and N. Gordon. *Sequential Monte Carlo in Practice*. New York: Springer, 2001.
- [13] A. Doucet, S. J. Godsill, and C. Andrieu. “On sequential Monte Carlo sampling methods for Bayesian filtering”. In: *Statistics and Computing* 10.3 (2000), pp. 197–208.
- [14] A. Doucet and A. M. Johansen. “A tutorial on particle filtering and smoothing: fifteen years later”. In: *Handbook of Nonlinear Filtering*. Ed. by D Crisan and B Rozovsky. Cambridge: Cambridge University Press, 2010.
- [15] R. Holenstein. “Particle Markov Chain Monte Carlo”. PhD thesis. Vancouver: University Of British Columbia, 2009.
- [16] N. Kantas, A. Doucet, S. S. Singh, and J. M. Maciejowski. “An overview of sequential Monte Carlo methods for parameter estimation in general state-space models”. In: *15th IFAC Symposium on System Identification*. 2009, pp. 774–785.
- [17] H. R. Künsch. “Particle filters”. In: *Bernoulli* 19.4 (Sept. 2013), pp. 1391–1403.
- [18] A. Lee and K. Latuszynski. “Variance bounding and geometric ergodicity of Markov chain Monte Carlo kernels for approximate Bayesian computation”. In: *Biometrika* 101.3 (2014), pp. 655–671.
- [19] F. Lindsten, R. Douc, and E. Moulines. “Uniform ergodicity of the Particle Gibbs sampler”. In: *arXiv.org* (Jan. 2014). arXiv:1401.0683v1 [math.ST].
- [20] J. S. Liu. “Metropolized independent sampling with comparisons to rejection sampling and importance sampling”. In: *Statistics and Computing* 6.2 (1996), pp. 113–119.
- [21] J. S. Liu and R. Chen. “Blind deconvolution via sequential imputations”. In: *Journal of the American Statistical Association* 90.430 (1995), pp. 567–576.
- [22] R. M. Neal. “Annealed importance sampling”. In: *Statistics and Computing* 11.2 (2001), pp. 125–139.
- [23] N. Whiteley, A. Lee, and K. Heine. “On the role of interaction in sequential Monte Carlo algorithms”. In: *arXiv.org* (Sept. 2013). arXiv:1309.2918v1 [stat.CO].

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

URL: <http://jhhuggins.org/>

E-mail address: jhuggins@mit.edu

UNIVERSITY OF TORONTO

URL: <http://danroy.org/>

E-mail address: droy@utstat.toronto.edu